BIOCHEMICAL AND STRUCTURAL PROPERTIES OF *POTATO VIRUS A* VPg

Kimmo Rantalainen

Department of Applied Biology
and
Department of Applied Chemistry and Microbiology
Faculty of Agriculture and Forestry
and
Viikki Graduate School in Biosciences

University of Helsinki

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of University of Helsinki, for public criticism in the auditorium Ls3 of building B (Latokartanonkaari 7, Helsinki), February 5th, 2010, at 12 o'clock noon.

Supervisor
Docent Kristiina Mäkinen
Department of Applied Chemistry and Microbiology
University of Helsinki
Finland

Reviewers
Professor Paavo Kinnunen
Institute of Biomedicine
University of Helsinki
Finland

And

Professor Erkki Truve
Department of Gene Technology
Tallinn University of Technology
Estonia

Opponent
Professor Peter Tompa
Institute of Enzymology
Hungarian Academy of Sciences
Hungary

Cover figure: Vesicle interaction and structural stabilization of PVA VPg.

ISBN 978-952-10-6013-7 (paperback)
ISBN 978-952-10-6014-4 (PDF)
ISSN 1795-7079

Helsinki University Print
Helsinki, Finland 2010

List of Original Publications

The following original publications are referred to using the bold roman numerals throughout the text.

I **Kimmo I. Rantalainen,** Vladimir N. Uversky, Perttu Permi, Nisse Kalkkinen, A. Keith Dunker, Kristiina Mäkinen. 2008. Potato virus A genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. Virology 377:280–288

II **Kimmo I. Rantalainen**, Katri Eskelin, Satu Hyvärinen and Kristiina Mäkinen. Investigating the functions of a lysine-rich region within viral genome-linked protein, VPg, of Potato virus A in replication and translation. Manuscript.

III **Kimmo I. Rantalainen**, Peter A. Christensen, Anders Hafrén, Daniel E. Otzen, Nisse Kalkkinen, Kristiina Mäkinen. 2009. Interaction of a potyviral VPg with anionic phospholipid vesicles. Virology. 395:114-120.

ABBREVIATIONS

α –MORE	α–helical molecular recognition element
ANS	1-anilino-8-naphthalenesulfonic acid
cAMP	Cyclic adenosine monophosphate
CD	Circular dichroism
CI	Cylindrical inclusion protein
CaMV	Cauliflower mosaic virus
DHN	Dehydrin
DNA	Deoxyribonucleic acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPG	1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
DOPS	1,2-dioleoyl-sn-glycero-3-phospho-L-serine
eIF4E	Eukaryotic initiation factor 4E
EM	Electron microscopy
ENSA	Endosulfine-α
ER	Endoplasmic reticulum
ERES	ER exit site
FHV	Flock house virus
FRET	Fluorescence resonance energy transfer
GdmHCl	Guanidinium hydrochloride
HA	Hemagglutinin
HCA	Hydrophobic cluster analysis
HcPro	Helper component proteinase
HPLC	High pressure liquid chromatography
IDP	Intrinsically disordered protein
IRES	Internal ribosomal entry site
MBP	Myelin basic protein
MP	Movement protein
mRNA	Messenger ribonucleic acid
NIa-Pro	Nuclear inclusion a-proteinase
NIb	Nuclear inclusion b-polymerase
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NTP	Nucleotide triphosphate
ORF	Open reading frame
PDB	Protein data bank
PIP	Phosphatidylinositol phosphate
PIPO	Pretty interesting potyviviridae ORF
PVA	Potato virus A
PVIP	Potyvirus VPg interacting protein
PVY	Potato virus Y
RdRp	RNA-dependent RNA Polymerase
RNA	Ribonucleic acid
SARS	Severe acute respiratory syndrome
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TFE	2,2,2-Trifluoroethanol
TMV	Tobacco mosaic virus
TuMV	Turnip mosaic virus
UTP	Uridine 5'-triphosphate
VPg	Viral genome linked protein

CONTENTS

LIST OF ORIGINAL PUBLICATIONS
ABBREVIATIONS
ABSTRACT **7**
INTRODUCTION **8**
1. Potyviruses **8**
2. Functions and modifications of viral genome-linked proteins (VPgs) **10**
2.1 Replication 10
2.2 Translation and translation inhibition 11
2.3 Nuclear and nucleolar localization and inclusions 12
2.4 Movement 13
2.5 Phosphorylation 13
3. Intrinsically disordered proteins **14**
3.1 Biochemical and structural properties of intrinsically disordered proteins 15
3.2 Functions of intrinsically disordered proteins 16
3.3 Structural stabilization of intrinsically disordered proteins 16
3.3.1 Membrane interactions associated with structural stabilization 17
3.4 How to study the lack of structure? – A methodological point of view. 18
3.4.1 Circular dichroism spectroscopy 18
3.4.2 Nuclear magnetic resonance 19
3.4.3 Fluorescence spectroscopy 20
3.4.4 Sodium dodecyl sulfate polyacrylamide gel electrophoresis, size exclusion chromatography, and limited proteolysis 20
3.4.5 Bioinformatics 21
AIMS OF THE STUDY **22**
MATERIALS AND METHODS **23**
RESULTS AND DISCUSSION **24**
1. Biochemical properties of PVA VPg **24**
1.1 PVA VPg in NTP–binding and uridylylation 24
1.2 PVA VPg in translation inhibition 25
2. Structural properties of PVA VPg **26**
2.1 The proportion of disorder in VPgs 26
2.2 Localization of the disordered regions 27
2.3 Stabilization of intrinsic disorder 30
2.4 Vesicle modifications caused by VPg – Evidence for a pore forming activity 30
2.5 Implications of membrane modifications for virus biology 34
CONCLUDING REMARKS **36**
ACKNOWLEDGEMENTS **37**
REFERENCES **38**
REPRINTS OF ORIGINAL PUBLICATIONS

ABSTRACT

The potato virus A (PVA) genome linked protein (VPg) is a multifunctional protein that takes part in vital infection cycle events such as replication and movement of the virus from cell to cell. VPg is attached to the 5´ end of the genome and is carried in the tip structure of the filamentous virus particle. VPg is also the last protein to be cleaved from the polyprotein. VPg interacts with several viral and host proteins and is phosphorylated at several positions. These features indicate a central role in virus epidemiology and a requirement for an efficient but flexible mechanism for switching between different functions.

This study examines some of the key VPg functions in more detail. Mutations in the positively charged region from Ala38 to Lys44 affected the NTP binding, uridylylation, and *in vitro* translation inhibition activities of VPg, whereas *in vivo* translation inhibition was not affected. Some of the data generated in this study implicated the structural flexibility of the protein in functional activities. VPg lacks a rigid structure, which could allow it to adapt conformationally to different functions as needed. A major finding of this study is that PVA VPg belongs to the class of ´intrinsically disordered proteins´ (IDPs). IDPs are a novel protein class that has helped to explain the observed lack of structure. The existence of IDPs clearly shows that proteins can be functional and adapt a native fold without a rigid structure.

Evidence for the intrinsic disorder of VPg was provided by CD spectroscopy, NMR, fluorescence spectroscopy, bioinformatic analysis, and limited proteolytic digestion. The structure of VPg resembles that of a molten globule-type protein and has a hydrophobic core domain. Approximately 50% of the protein is disordered and an α-helical stabilization of these regions has been hypothesized. Surprisingly, VPg structure was stabilized in the presence of anionic lipid vesicles.

The stabilization was accompanied by a change in VPg structure and major morphological modifications of the vesicles, including a pronounced increase in the size and appearance of pore or plaque like formations on the vesicle surface. The most likely scenario seems to be an α-helical stabilization of VPg which induces formation of a pore or channel-like structure on the vesicle surface. The size increase is probably due to fusion or swelling of the vesicles. The latter hypothesis is supported by the evident disruption of the vesicles after prolonged incubation with VPg. A model describing the results is presented and discussed in relation to other known properties of the protein.

INTRODUCTION

1. Potyviruses

The genus *Potyvirus* is a group of positive–sense RNA plant viruses belonging to the virus family *Potyviridae*. *Potyvirus* is a significant genus in the distributional, economic, and scientific sense. The type species of the genus is *potato virus Y* (PVY), a close relative of *potato virus A* (PVA). Evolutionarily, the birth of the genus is currently thought to have followed the emergence of agriculture approximately 6,600 years ago [1]. This recent emergence together with rapid spread and speciation go hand in hand with the agricultural achievements of humans.

The major features of the potyvirus infection cycle have been outlined. Aphids are the insect vector in which the virus occurs in a non-persistent manner [2]. An aphid injects the virus through its stylus into the host plant cell. After cell entry, the virus particle is disassembled and the polyprotein is translated. Translated proteins assemble the replication machinery which produces copies of the genetic material for viral progeny. As more viral proteins are translated, the viral RNA is packaged inside particles for cell-to-cell transmission, long distance dissemination in the host plant, or plant-to-plant movement by aphid transmission [3]. This outline, however, is an oversimplification of the virus infection cycle events, and at the biochemical

Table 1. *Known functions of potyvirus proteins. VPg is clearly among the most versatile potyviral proteins.*

Protein	Functions	Reference
P1	Proteinase, RNA silencing suppression	[192-194]
HcPro	Aphid transmission, proteinase, cell-to-cell and long-distance movement, gene silencing suppression	[74, 195-198]
P3	Movement, replication.	[199, 200]
6K1	Cell-to-cell movement	[201]
CI	RNA helicase, ATPase, RNA binding, movement	[71, 202]
6K2	Membrane anchor, long-distance movement	[185,186,203]
VPg	Replication, NTP binding, RNA binding, cell-to-cell and long distance movement, translation inhibition, gene silencing suppression	[16, 18, 56, 59, 66, 75, 169, 204]
NIa-Pro	Protease, DNase	[205, 206]
NIb	RNA dependent RNA polymerase (RdRp), replication, uridylylation	[16, 207, 208]
CP	Aphid transmission, cell-to-cell and long-distance movement, Encapsidation of viral RNA, Regulation of RNA amplification	[209-211]
PIPO	Unknown	[4]

level, we realize that there are a vast amount of different activities taking place.

Potyviruses are encoded by a ~10,000 nucleotide long (+)–stranded RNA genome. Potyviral genomes were thought to encode a polyprotein from a single open reading frame that is processed to ten mature proteins. Recently, however, a bioinformatic approach revealed that an 11th protein is translated from a +2 frame [4]. This protein, whose function is so far unknown, was named PIPO for Pretty Interesting Potyviridae ORF. Currently known functions of potyviral proteins are listed in Table 1. The virus coordinates the infection cycle events with this relatively low number of proteins. As we see from Table 1, many of the virus proteins are multifunctional and a prime example of such is the subject of this study – the viral genome linked protein (VPg). In Figure 1, VPg functions are presented in relation to the progression of the infection cycle inside the cell. Some of the roles presented are hypothetical and await a better understanding of the infection cycle details. To understand the functions of VPg, the key biochemical events and structural properties of VPg were examined in detail in this study.

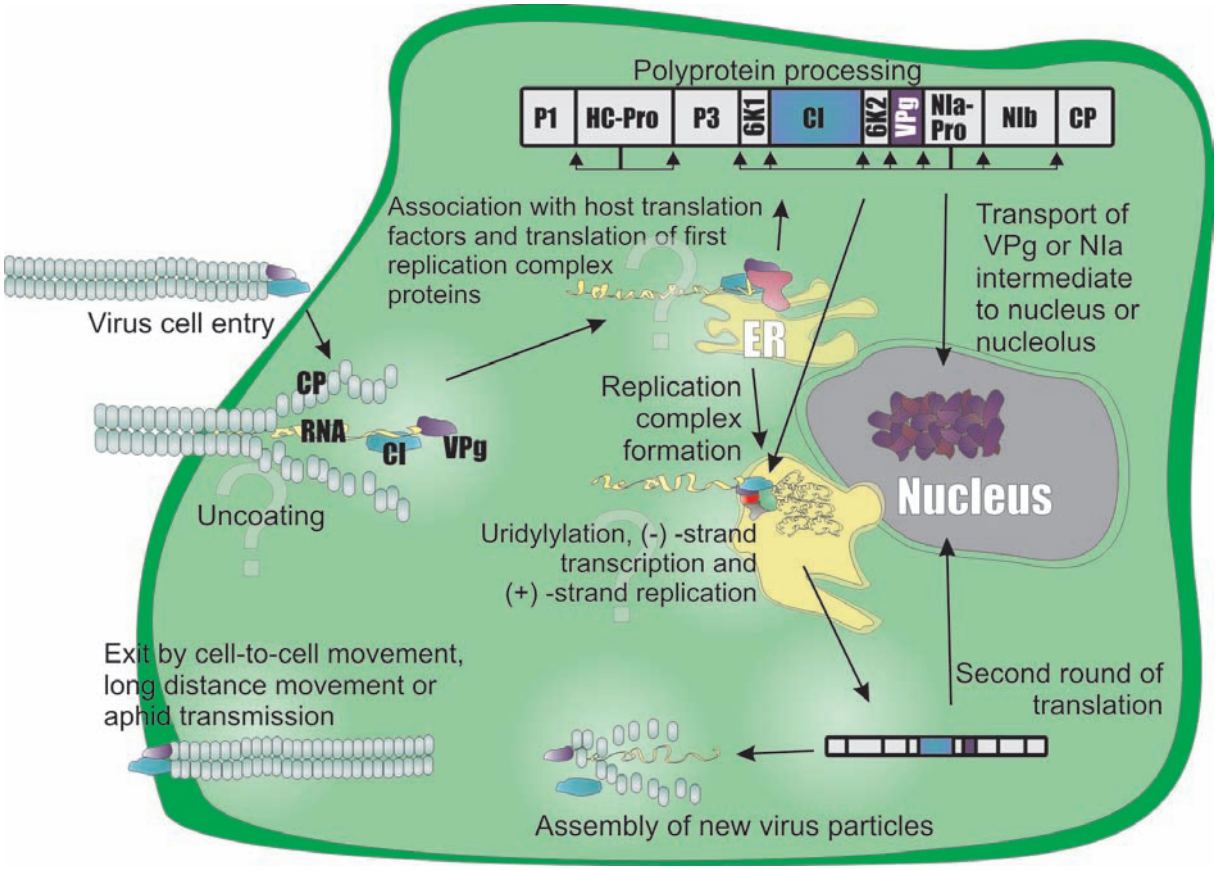


Figure 1. *Hypothetical infection cycle progress from the VPg point of view. Question marks indicate the most important phases in which the role of VPg is still poorly understood. Perhaps the most important open question is the formation of the replication complex. For example, it is unknown whether the initial round of translation, replication and subsequent rounds of translation occur within the same membrane environment. The membrane environment, however, is clearly required for both activities.*

2. Functions and modifications of viral genome-linked proteins (VPgs)

The current understanding of the molecular level events in the life cycle of plant viruses has been built mostly during the last 20 years. Technological advances have made it possible to study, for example, direct protein-protein interactions and localizations of virus proteins inside host plant cells. These advances have also brought the current understanding of VPg proteins to a new level. It is not a single function that stands out from this knowledge, but the multifunctionality of the VPg protein. VPg takes part in replication, translation inhibition, and virus particle movements (see Table 1 for references). How then is the regulation between different functions enabled? This thesis presents the interesting and novel property of instrinsic protein disorder as a partial answer to this question. Several plant viral VPgs seem to be intrinsically disordered proteins (IDPs) and, in fact, examples from many different protein families from all kingdoms of life have recently been described [5-7]. This class of proteins is presented in more detail later in this study.

An additional complexity in the multifunctional and unstructured VPg functions is introduced by the viral polyprotein processing intermediates. The intermediates can interact with host factors, but little is known regarding the functions of these interactions [8]. The polyprotein intermediate of VPg is associated with *Tobacco etch virus* virions in addition to the fully processed VPg, whereas in PVA only the fully processed VPg is detected [9, 10]. The NIa-Pro – VPg intermediate is the last site to be cleaved in PVA polyprotein processing [11]. Fully processed VPg has a molecular size of 22 kDa and can interact with host factors and other viral proteins [12-14] as well as forming dimers and, occasionally, higher oligomers. The functional purpose of multimer formation remains unknown [6, 10]. One study has suggested that dimerization could be an artifact originating from a recombinant protein purification protocol [6].

The study of protein-protein interactions and especially structural changes in proteins are difficult or impossible *in planta*. The closest *in vivo* technique has long been the yeast two-hybrid system, which has provided important background information and evidence of viral protein-protein interactions [12, 13]. Recently, bifluorescence based systems (e.g. BiFC and FRET) have started to yield valuable information on protein interactions, as well as the localization of the interaction [15]. However, most protein-protein interaction studies rely on *in vitro* methods, and biophysical protein structure studies are almost exclusively performed *in vitro*.

2.1 Replication

Potyviral replication most probably starts with viral polymerase (NIb) driven uridylylation [16, 17]. Uridylylation results in initiation of new RNA synthesis, in which VPg acts in a non-enzymatic manner as the primer. Uridylylated VPg is thought to act as the primer for both (-) and (+) strands of the genome [2].

Uridylylation of PVA VPg was first described by Puustinen and Mäkinen [16]. A lysine rich nucleotide binding site of VPg was also described. This site most probably guides the approaching UTP nucleotide to the exact site of uridylylation. The potyviral uridylylation reaction itself is targeted to the conserved tyrosine residue corresponding to Tyr63 of the PVA VPg (starting from the 5' of the VPg amino acid sequence) [18-20]. A more detailed mechanism for uridylylation has been presented by Schein *et. al.*, based on the NMR structure of the poliovirus VPg [21,22]. Although there are major differences between potyviral and picornaviral VPgs,

including a considerable difference in size, the lysines of the NTP binding site appear to be conserved and may participate in the reaction through a similar mechanism [16]. In the poliovirus model, the uridylylation reaction is coordinated by a magnesium ion and the charged lysine and arginine side chains. This arrangement allows the nucleotide to approach the tyrosine in the correct orientation and the viral polymerase to attach the nucleotide to the VPg. This reaction results in a covalent link between the α–phosphate of UTP and the tyrosine OH group of VPg and, consequently, the initiation of new RNA strand elongation according to the template strand nucleotide sequence. Other components are probably required for replication to proceed, such as the host poly(A) binding protein (PABP), which interacts directly with the potyviral polymerase [23, 24]. It seems, however, that the initiation of potyviral replication is template independent whereas the VPgs of many other families, such as picornaviruses, require the template [16, 25].

What else is needed for virus genome replication to occur? Generally, the (+) sense RNA virus replication complex is composed of template, newly synthesized RNA, replicase, and host factors. The replicase can be composed of RdRp, a helicase, and a methyltransferase activity [2]. In potyviruses, the methyltransferase activity is not required since the RNA genome binds VPg at the 5' end in place of a 5' methyl cap. The number of replication associated viral proteins can vary. For potyviruses, at least RdRp, VPg, and CI are required, whereas replication factor protein A alone is sufficient to provide all activities needed for replication in the case of the *Flock house virus* (FHV) [26]. Replication of (+) strand RNA viruses requires a (-) strand RNA intermediate. However before the (-) strand can be transcribed, an initial round of translation must occur to produce the replication associated viral proteins, as presented in Fig. 1. Therefore, the replication complex assembly follows the unpacking of the entering virus particle and the translation of the first viral proteins. Before replication can start, the complex needs to acquire membranous components to create a suitable compartment, as well as appropriate conditions and support for the multiplication of the RNA [27, 28].

The membranous environment is an absolute requirement for the assembly of a functional replication complex in (+) sense RNA viruses (reviewed in [28-30]). In addition to the above mentioned FHV, examples include *Poliovirus, Semliki forest virus, Hepatitis C virus, Tomato ringspot virus,* and many others [26, 31-36]. At least six different plant organelles are associated with (+) stranded plant virus replication [2]. The membranous environment may be derived from organelles such as endoplasmic reticulum (ER), mitochondria, chloroplasts, vacuole, or peroxisomes [37]. The types of modifications made to these organelles ranges from sacks or miniorganelles to pore complexes and inclusion bodies. Replication associated proteins have been shown, for example, to form inclusions on the surface of *Tobacco mosaic virus* (TMV) infected plant cell ER membranes [38]. The exact mechanisms by which these modifications work vary. But, the common end result is that the RNA is replicated, and the newly produced strand and the VPg, if such is included in the virus proteome, are ready to move forward to the next stage of the infection cycle.

2.2 Translation and translation inhibition

Mature eukaryotic cell transcripts are 5' capped mRNAs that are transported out of the nucleus into the cytoplasm for protein translation. The cap structure acts as a translation machinery assembly signal. Eukaryotic translation initiation factor 4E (eIF4E) plays a central role in recognizing the cap structure, and initiating the assembly of the initiation complex [39]. eIF4E interacts directly with VPg. For potyviruses, the eIF4E – VPg interaction was first demonstrated in

the *Turnip mosaic virus* (TuMV) and has since been a major point of interest in potyvirus studies [15, 40-47].

Eventually, the translation machinery includes several other components, such as other translation initiation factors, poly(A) – binding proteins, and ribosomal subunits, and starts to synthetisize the polypeptide chain [48, 49]. Viruses need to trick this machinery into producing the viral proteins and VPg could play a role in that process [50, 51]. Some viruses encode their own methyltransferase activity for capping their RNA or an internal ribosomal entry site (IRES), which is suggested to be a structural feature of the mRNA downstream of the 5′ end that is recognized by eukaryotic ribosomes. VPg may act as a protein substitute for a methyl-cap or IRES. However, there is no direct evidence for VPg participation in translation and cap independent translation is possible for viruses that carry VPg in their genome even if VPg is absent [52, 53]. In contrast, there is *in vitro* evidence for the translation inhibition activity of VPg [45, 54-56]. This activity may be part of the viral protein translation induction or it may be attributed to a different activity. One interesting possibility is that the nuclear localization of VPg leads to a disruption of host mRNA transport and eventually to a decrease in host protein translation. The decrease in host protein translation could also occur by inhibition of complex formation between cellular host mRNAs and eIF4E, as suggested by Miyoshi *et al.* [56].

As presented in Fig. 1, potyvirus protein translation occurs in two different stages. First, an initial round is needed to produce the viral components for the replication complex. In the second stage, the replicated viral genome is used as a template for translation and larger amounts of viral proteins needed for virus particle assembly are produced. As discussed above, the precise processes in which VPg takes part are currently unclear, but given the multiplicity of the activities taking place and the subsequent requirement for regulation, it is possible that VPg functions in several stages in different cellular compartments and it may even possess the activity to switch, *e.g.* from transcription to translation, and hence be the regulator itself. Recently, cap-bound eIF4E has been shown to bind VPg [47, 57], which would enable it to take part in both of the activities simultaneously.

2.3 Nuclear and nucleolar localization and inclusions

Potyviral VPg contains two nuclear localization signals (NLS) which were mapped to the positively charged N-terminus of TEV and PVA [58, 59]. In PVA, the first NLS spans from Lys4 to Lys6 and the second from Lys41 to His55. The first NLS was also shown to target the NIa intermediate to the nucleolus and to Cajal bodies. The second NLS signal overlaps with the NTP binding site reported earlier to span from Ala38 to Lys44. Deletion of this region also disrupts uridylylation [16]. This positively charged region is undoubtedly important in at least these two functions and possibly contributes to other interactions and the conformation of the protein with its high net positive charge. The functions of potyvirus VPg inside the plant cell nucleus remain poorly understood, but there are some observations that suggest that VPg could take part to gene silencing suppression. For example, the 2b protein of the *Cucumber mosaic virus*, a member of the *Bromoviridae* family, suppresses post-transcriptional gene silencing [60] and deletion of NLS affects the gene silencing activity of PVA VPg [59]. The movement protein of a *Barley yellow dwarf virus* from the genus *Luteovirus* contains a NLS which was shown to target the protein to the nuclear envelope and to form amphiphilic α–helices on the surface of the nuclear envelope [61, 62]. The function of the protein at the envelope surface has been suggested to be RNA transport across the membrane bilayer. Several animal virus proteins have been shown to hamper the nuclear functions

of the host and to bind to viral RNA inside the nucleus and affect the replication of the viral RNA [63].

Several viral proteins have been termed 'inclusion' proteins, since they were initially discovered as inclusions [64]; for example, the nuclear inclusion protein A (NIa), which is a polyprotein intermediate containing VPg in the N–terminal region and a proteinase in the C–terminal region. Inclusions are sometimes considered to be waste depositories of excess virus proteins produced from the polyprotein. There is, however, a considerable amount of evidence suggesting that inclusions may have a function of their own. One example is the role of inclusions as sites of translation. TMV replicase protein inclusions have functional importance and were localized to the ER surface [65]. In the light of recent data, it would appear that for potyviruses there is distinct localization of VPg/Nla proteins to the nucleolus and Cajal bodies, which also puts into question the presence of nuclear inclusions as waste depositories [59].

2.4 Movement

Movement of plant viruses from cell-to-cell or to more distant locations within the plant are usually mediated by a group of proteins called movement proteins (MPs). Potyviruses use multiple proteins, including VPg, instead of MPs [66]. The exact mechanisms of VPg in movement related functions are largely unknown. Interactions between host and viral factors seem to be more important than the intrinsic features of VPgs. The interaction with the descriptively named host factor, potyvirus VPg-interacting protein (PVIP), and the host factor eIF4E are thought to mediate VPg related cell-to-cell movement functions [67, 68]. The cylindrical inclusion protein (CI) has also been reported to take part in cell-to-cell movement via plasmodesmata [69-71]. An atomic force microscopy study revealed a tip structure at the end of the PVA particle which contained the CI, VPg, and the helper component–proteinase (HcPro) [72, 73]. The function of this structure in virus movement could be in forming a structure for inserting the viral genome inside the host cell. This insertion could occur through plasmodesmata since both HcPro and the coat protein can mediate cell-to-cell movement by increasing the size exclusion limit of the plasmodesmata [74]. The role of the potyviral VPg in long-distance movements may be related to phloem loading and translocation to the infected leaves in the early stages of systemic infections [75, 76].

2.5 Phosphorylation

Phosphorylation is a common mechanism to regulate protein function and structure. The PVA VPg is phosphorylated both *in vitro* and *in vivo* at multiple sites [77-79]. Phosphorylation supplies an important level of regulation, although the exact functions of the phosphorylations are unknown for potyvirus VPgs. Phosphorylation could trigger the disassembly of the virus particle upon entering the cell [78], having similar effect compared to phosphorylation of the potato virus X coat proteins (*Potexvirus* genus) and the TMV MPs (*Tobamovirus* genus) [80-82]. More specifically, TMV MP initial phosphorylation was shown to affect plasmodesmal permeability and further phosphorylation was hypothesized to inactivate movement function and trigger disassembly [83].

Phosphorylation and other posttranslational modifications are important for structural regulation of intrinsically disordered proteins. A bioinformatic study with DISPHOS, a prediction software that can predict phosphorylated amino acids in disordered regions with 76 – 83% accuracy, depending on the phosphorylated amino acid, supports the hypothesis that protein phosphorylation occurs predominantly within the intrinsically disordered regions [84]. The best known example of an IDP in which phosphorylation serves as a structure regulator is the myelin basic protein (MBP)

of central nervous systems [85, 86]. MBP is a multifunctional protein interacting with negatively charged lipid layers, actin, and tubulin, as well as with itself when forming dimers, and contains a nuclear localization signal. It has a net positive charge which is decreased by phosphorylation. This decrease affects the binding activities of the protein by weakening the electrostatic interactions and, thus, reveals a precise mechanism for functional regulation [87]. MBP appears to have similar biochemical properties compared to PVA VPg but whether a similar type of regulation mechanism by phosphorylation holds true for VPgs, remains to be shown.

3. Intrinsically disordered proteins

The concept of intrinsically disordered proteins challenges one of the basic principles of protein structure – function relationships. With IDPs, we can no longer consider protein function to rely on a rigid structure. Intrinsic disorder can be seen as the epigenetics of proteins; the function of the protein is not simply dictated by its sequence but also by its immediate surroundings, just as gene expression is regulated by epigenetic factors that lie outside the DNA sequence. The adaptive and flexible regions of an IDP can be also compared to the regulatory parts of genes such as promotors and other non-coding regions of genomes.

An increasing number of proteins have been identified to be intrinsically disordered, many of which were discovered when their structures could not be resolved after several trials. Most of the proteins that have been verified to be at least partly disordered are recorded in the DisProt database, which currently holds over 500 proteins [88]. It seems that the fading number of proteins proven to have rigid structures has exploded to yield a flood of intrinsically disordered proteins. Sequence analysis predicts that more than 15,000 of the proteins deposited in the Swiss Protein Database could contain a disordered region of at least 40 amino acids [89]. Furthermore, only 7% of the protein structures deposited in the Protein Data bank (PDB) cover the sequence of the whole protein [90]. In these cases, the missing amino acids could be part of an intrinsically disordered region.

Although these observations have not yet reached textbooks in biochemistry, they have generated a number of excellent reviews [91-99]. The authors of these reviews have had a major influence on the development of this fascinating branch of structural biology. Perhaps because of its novelty, there are still discrepancies in the nomenclature and IDPs are sometimes called ´natively unfolded´ or ´intrinsically unstructured´. The term ´intrinsically disordered´ is, however, suggested to be appropriate as a more general term describing all types of incompletely folded proteins and regions [97]. At the same time, the terms natively unfolded or intrinsically unstructured may be used to refer to subsets of IDPs that have little or no ordered structure and behave as random coils or pre-molten globules.

Relatively few plant virus proteins have been shown to be intrinsically disordered. The measles virus nucleoprotein has an N-terminal domain that folds into a α-helical conformation upon interaction with the viral polymerase cofactor phosphoprotein (P) [100]. Other examples include the *Hepatitis C virus* core protein and the *Bovine viral diarrhea virus* core protein [101, 102].

Intrinsically disordered proteins have been suggested to play a role in the virulence of influenza virus strains [103]. Increased virulence was correlated with increased disorder in the hemagglutinin (HA) protein, which is an influenza virus surface protein that mediates viral cell entry. Recent reports have shown that plant viral VPgs also belong to this class of proteins [5-7]. Intrinsic disorder offers a way to understand multifunctionality and gives a starting point for structural studies of the protein. The following sections present the IDP class in more detail. Because the identification of intrinsic protein disorder usually requires a combination of several biochemical and biophysical methods, a methodological approach is included.

3.1 Biochemical and structural properties of intrinsically disordered proteins

IDPs lack a rigid structure under physiological conditions and in their native fold. Instead, they have an ensemble of folds rapidly changing from one to another. IDPs can, however, have a bias towards a certain fold which is dictated by the prevailing conditions and structure regulating factors such as posttranslational modifications. Because IDPs are in constant flux between different structures, IDPs can be considered transiently structured. In this sense, the differences between IDPs and stably structured proteins are plainly biophysical. The positions of atoms and the amino acid chain backbone angles of stable proteins have equilibrium positions whereas, in disordered proteins, these values vary considerably. However, as the difference in the degree of variability in these values between a structured and disordered protein is not defined, intrinsic disorder remains a qualitative description [91,104]. The structure of an IDP has also been suggested to resemble the denatured state of structured or stable proteins [98, 99]. Therefore, an important question is what is the difference between an intrinsically disordered protein and the denatured state of a structured protein? It is not the degree of folding, since denatured proteins can still have some structured elements as do IDPs. The most significant difference is the biological activity, which is present in intrinsically disordered proteins, but not in denatured proteins.

Compared to structured proteins there is also a clear difference in amino acid composition. Disordered proteins have a significant enrichment of Ala, Arg, Gly, Gln, Ser, Pro, Glu and Lys residues, and reduced levels of Trp, Cys, Phe, Ile, Tyr, Val, Leu and Asn residues [105, 106]. The length of the disordered region can be from a few amino acids to long stretches or even cover the whole protein. The enrichment of certain amino acids can lead to distinct secondary properties such as a high isoelectric point (pI).

IDPs can be divided into two subclasses by their secondary structure content. The members of the coil-like (or intrinsic coils) group are less compact and behave as random coils. The molten globule-like group, also called natively unfolded proteins, consists of proteins which are more compact and typically exhibit some degree of ordered structure [99]. The most convinient way to determine which class the protein of interest belongs to is to determine the circular dichroism (CD) spectrum of the protein in the far UV region. Coil-like proteins have closer to zero ellipticity (~ -2,000 deg/cm^2/dmol^{-1}) at 222 nm and lower ellipticity at 200 nm (~ -20,000 deg/cm^2/dmol^{-1}) compared to molten globule-like IDPs. A common feature of both groups is the loose packing that evidently leads to a larger hydrodynamic radius and larger interaction surface. This is an advantage when the protein has multiple interaction partners as it increases the interaction area and structural adaptation capability. Some disordered regions don't function by adapting a specific structure themselves but serve to increase the mobility of other parts of the protein. Such regions are usually called loops, linkers, or hinges. A single protein can have several unstructured

and structured regions that can interact with each other and affect the overall folding. This type of protein, such as the extensively studied tumor suppressor p53 [107], presents a difficult challenge for the structural biologists but also offers a way to understand the interplay between protein structure and fundamental biochemical functions.

3.2 Functions of intrinsically disordered proteins

Structural flexibility and functional diversity seem to go hand in hand for intrinsically disordered proteins [108]. An interesting concept presented to clarify the reasoning for this phenomenon is the theory of flexible nets and hub proteins [108, 109]. This hypothesis suggests that the intrinsically disordered protein is the hub controlling a flexible network of interactions and functions. When the hub protein interacts with its binding partner it leads to structural adaptation of the IDP. This adaptation does not necessarily lead to stabilization of the whole protein since the region gaining more structure may be a separate domain or a small interaction surface region. As mentioned, MBP is an example of an IDP whose structure is regulated by phosphorylation. It is also a protein that has been suggested to act as a hub for structural and signaling networks [110]. Another well studied example of a protein that could be considered a hub protein is the tumor suppressor p53. Oldfield *et al*. has carefully reviewed the solved structures and binding partners of p53 and pointed out that this protein has multiple disordered regions which enable it to interact it with multiple partners and to adopt several different structures. This example also emphasizes the central role of disordered hub proteins. If p53 fails to do its work, *e.g*. because of a mutation, the well studied consequences include the development of several cancer types due to failure to maintain the appropriate signaling network [111].

What, then, are the functions that are activated by the interactions with or modifications to an IDP? There seems to be no single obvious functional category where IDPs would fall. Rather, the structural flexibility appears to be related to many different types of functions from all kingdoms of life [97, 104, 112-115]. There are, however, some functional categories where intrinsic disorder is typically found, the most common being related to DNA or RNA binding, signal transduction, ribosomes, and membrane binding. More specifically, the DNA and RNA binding functions seem to be related to transcriptional and translational regulation. Several transcription factors such as *Arabidopsis* HY5 are partly disordered [116]. In fact, a bioinformatic study suggested that ~90% of transcription factors may possess extended regions of intrinsic disorder [117]. In contrast, the names of ordered proteins often seem to end with “ase” [97]. Therefore, enzymatic activity is probably associated with a more stable protein structure.

3.3 Structural stabilization of intrinsically disordered proteins

Interactions and functional activation associated structural stabilization of IDPs can occur through several different mechanisms. A single IPD region can be stabilized or a single binding event can stabilize several regions. Another possibility is that the unstructured domains are stabilized sequentially, one region being dependent on the other. Post-translational modifications bring yet another level of regulation to the structural stabilization cascade. As a reversible reaction, a post-translational modification can turn the activity of the protein on or off by inducing a disorder to order, or an order to disorder, transition [118]. Wright and Dyson proposed a division of structural stabilization into two extreme classes [94]. The first class is the induced folding mechanism where the disordered protein adapts its structure to a binding partner. The second class is the conformational selection

method, where the binding partner selects the best conformational counterpart from the ensemble of disordered conformational intermediates [94]. Examples of the first mechanism, induced folding, are the acidic transcriptional activators Gal4 and VP16 which are recruited through electrostatic interactions to the target proteins leading to formation of a target induced and stabilized complex [119]. PEP-19, a small calmodulin–binding protein, is an IDP which is suggested to bind to either the apo- or Ca^{2+}- form of calmodulin with different structural modes and, therefore, represents the conformational selection method of structural stabilization [120]. A more precise mechanism for the induced folding is described by the concept of α–helical molecular recognition elements (α–MOREs) [121]. α–MOREs are short regions within longer disordered regions which undergo coupled folding and binding upon interaction with a binding partner. The induced folding of the measles virus nucleoprotein C-terminal domain for example uses this type of recognition mechanism when binding to the viral phosphoprotein [100].

Biological reactions happen within the laws of thermodynamics and, therefore, disorder in a protein must be thermodynamically favorable. From this perspective, the rationale for intrinsic protein disorder can be based on stereochemical and thermodynamic aspects of the folding mechanisms. These are covered in the "fly-casting" theory [122], which is partly parallel to the induced folding concept and provides a kinetic description of how the loosely packed interaction surface of an IDP folds through a funnel-like channel towards the interaction, binding and stabilization. A critical assessment of the fly-casting mechanism recently added that, in addition to a greater capture radius of the IDP, there is also a significant reduction in the binding free-energy barrier [123]. This theory was applied successfully, for example, for deciphering the kinetics of the binding of the disordered CREB transcription factor pKID domain to the KIX domain of the co-activator CBP [124].

The question of coupled folding and binding, induced folding and conformational selection of IDPs is clearly open and under debate. The above mentioned critical assessment of the fly-casting mechanism is part of this debate. Additionally, another recent review of the induced folding and the conformational selection mechanisms argued that the two mechanisms are actually the same and a synergistic model was proposed based on the view that coupled folding and binding follow conformational selection [125].

3.3.1 Membrane interactions associated with structural stabilization

According to a bioinformatic study, intrinsically unstructured segments appeared to be significantly enriched in transmembrane proteins and, in approximately 70% of these, the unstructured segments were close to the transmembrane segments [126]. Since we know that positive–sense plant viruses need membranes for crucial steps in the infection cycle and that they induce modifications in membrane morphology [26-30, 127], it is essential to review some of the key aspects of membrane-IDP interactions. These interactions are presented here with a few well known examples.

α-synuclein is an important and well characterized IDP. It is also known to interact with acidic membrane lipids and cause a leakage of the vesicular contents through pore-like structures [128, 129]. The stabilization was shown to be accompanied by a 3% to 80% increase in α–helical content. More recently, the binding was shown to be dependent on the vesicle size, lipid/protein ratio, temperature, ionic strength, and lipid charge [130]. Endosulfine-α (ENSA) is a cAMP regulated phosphoprotein and a binding partner of α-synuclein. ENSA is also an IDP whose structure is stabilized upon membrane binding. The interaction between ENSA and α-synuclein may be

involved in regulating neuronal excitability and synaptic vesicle release in a membrane-dependent fashion [131]. Across the different kingdoms of life, α–helical stabilization is clearly a common way to modify structure upon interaction with a membrane. Maize dehydrins (DHNs) use a similar type of membrane stabilization mechanism. They have a positively charged "K segment", which interacts with anionic phospholipid vesicles and stabilizes into α–helices. DHNs accumulate in seeds and vegetative tissues of maize in response to abiotic stress conditions and during the late stages of embryogenesis. Structural adaptation of DHN proteins are presumed to be associated with the protection of the cellular components from stress damage [132,133]. These examples show that the membrane binding associated structural adaptation of an IDP can be a fine tuned and precise mechanism, and can lead to dramatic consequences.

3.4 How to study the lack of structure? – A methodological point of view.

X-ray crystallography and NMR are powerful methods to resolve the structure of a stable protein. Solved structures are usually thought to cover the entire protein, but a significant amount of solved structures contain "ambiguous" regions of unknown structure [90]. Therefore, there must be something that can not be seen with these methods and is in constant dynamic fluctuation. In order to understand the function of an IDP, it is essential to find ways to characterize its structure.

How then can we study the absence of something? There is no single best way to identify and characterize the structure of an IDP but, on the contrary, the best way is to combine several methods. It may be possible to stabilize the structure and use conventional structure resolving methods, but it may also be possible to observe the ensemble of structures and try to study the properties of the moving (disordered) parts by using different approaches. Several thorough reviews are available regarding the methodology used to identify an IDP [91,134-137]. The following chapters present more details on the nature of IDPs, as well as some model IDPs from a methodological point of view. The emphasis is on the methods that were employed in the original publications related to this study (**I** and **III**), where more detailed descriptions of these methodological setups can be found. The methods described here are not the only possible approaches, but are perhaps among the most commonly used. CD–spectroscopy is frequently the first wet lab step and a very versatile one.

3.4.1 Circular dichroism spectroscopy

Perhaps the most widely used method for identifying and characterizing an intrinsically disordered protein is circular dichroism (CD) spectroscopy. It is based on the refraction of circularly polarized light from asymmetrical molecules [138, 139]. All proteins are asymmetrical because of the chirality of the peptide bond. The refraction of light is measured and usually collected as milli-degrees of light angle rotation. If the subject of the study is a protein, after background subtraction, the data is normalized on the basis of the number of amino acid residues, the protein mass, cuvette path length and protein concentration. After normalization, the data is in the form of $deg/cm^2/dmol^{-1}$ and is usually plotted against the wavelength spectrum. A spectrum in the 250 to 350 nm range is called the "near-UV" region. Signals in this region describe certain aspects of tertiary structure such as the environment of the aromatic amino acids and disulfide bonds [140]. The signals of this region may be overlapping and difficult to interpret but can give a useful fingerprint of the tertiary structure, or show the absence of it. Therefore, the overall use of CD spectra from the near-UV region in characterizing intrinsic disorder is rather limited. However, a spectrum in the range of 180 to 250 nm, referred as "far–

UV", may be very helpful. Spectra in this range are commonly used to characterize the secondary structure of proteins [140, 141]. The spectral properties of certain types of secondary structure elements are defined in the literature [138]. For example the α–helix gives rise to a characteristic positive peak at 192 nm and double minima at 208 and 222 nm. The typical far-UV CD spectrum characteristics of an intrinsically disordered protein are a large negative ellipticity near 200 nm, a negligible ellipticity near 222 nm, and an ellipticity near zero at 185 nm [135]. CD spectroscopy also enables the separation of disordered proteins into two classes, as described earlier. When the CD spectroscopy data is used for estimating the proportions of different secondary structure elements, it is fed into software developed for the purpose. These are typically neural network based algorithms which are trained with datasets of known structures and deconvolute the data to percentages of different secondary structure elements. Two examples of this type of software are K2d developed by Andrade *et. al*. [142] and CDPro by Sreerama and Woody [143], consisting of three parallel algorithms SELCON3, CDSSTR, and CONTIN.

The use of CD spectroscopy in studying intrinsic disorder can be extended by observing changes in the protein structure. The change can be induced, for example, by addition of a denaturant or an interacting partner. It is also possible to observe the effect of heating on protein stability and conformational changes by adding a precise heat control element to the CD spectroscope setup. Intrinsically disordered proteins have known behaviors under certain conditions which can be studied with these extensions. For example, they are not drastically affected by heat denaturation. Changes occur at relatively low temperature and are many times reversible [135].

These features make CD spectroscopy an excellent tool for characterizing intrinsically disordered proteins. There is, however, a major drawback. CD spectroscopy describes only the content of structural elements, not the location of them. For this purpose, other parallel methods have to be used and nuclear magnetic resonance (NMR) is the most powerful of them.

3.4.2 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) depends on the magnetic properties of the nuclei of atoms which can be measured in a controlled magnetic field. The NMR spectra consist of peaks or resonance signals caused by the relaxation of the nuclei. If the quality of the spectrum is adequate, it can be used to assign the amino acids of the protein and eventually to resolve the exact three dimensional folding of the protein. The fact that NMR studies can be conducted using proteins in solution separates it from the other powerful structural biology method, x-ray crystallography, which requires that the protein be in the form of a symmetrical crystal. Thus, NMR enables one to use proteins with fluctuating structures rather than a single rigid structure, which is a perquisite for decent crystal formation. Commonly, the first indication of an intrinsically disordered protein in an NMR spectrum is the poor dispersion of the resonance signals.

Several variations in NMR methodology have been developed and when carefully selected and used can give valuable information on disordered proteins [134, 144]. Perhaps the most promising new application is in-cell NMR spectroscopy which is one of the few methodologies for conducting protein structural studies *in vivo* [145, 146]. This application has already provided evidence that α-synuclein – a well known IDP – retains its intrinsic disorder inside the cell in molecularly crowded conditions [147]. Molecular crowding prevents the conformational change in α-synuclein that is detected in the more dilute conditions used in *in vitro* assays. NMR applications designed for deciphering the folding pathways of proteins

have also been successfully used for IDPs, as in the case of calpastatin folding studies. Calpastatin peptides were shown to fold into α–helices in a Ca^{2+} dependent manner [148, 149]. Another important example of a successful NMR study of protein folding is the folding of the pKID domain of the CREB transcription factor upon its interaction with the KIX domain of the co-activator CBP. Two NMR techniques were used: titration NMR and ^{15}N relaxation NMR. These approaches revealed that the interaction is mediated by an ensemble of transient encounter complexes stabilized by non-specific hydrophobic contacts [150]. This study is a prime example of the detail NMR is capable of resolving in dynamic structural environments. NMR techniques are probably the most powerful individual method for characterizing IDPs and their folding, but are also, perhaps, the most difficult to conduct and require the most sophisticated instrumentation.

3.4.3 Fluorescence spectroscopy

Fluorescence spectroscopy offers a variety of methods for studying intrinsically disordered proteins. The basic principle is to excite the sample with a certain wavelength of light and to follow the emission intensity in a particular area of the spectrum. In protein fluorescence, the emission spectrum is dependent on the aromatic amino acids in a coordinated environment. Certain molecules, such as iodide or O_2 can be used to "quench" the emitted fluorescence [151]. The rate and efficiency of the quenching depends on the conformation and exposure status of the amino acids. A neutral quencher, acrylamide, can interact with exposed and buried tryptophan residues. If the quenching of a folded protein is compared to one with denaturants added, the level of stable structure, or the lack of it, can be estimated. The quenching effect is visualized as a Stern-Volmer plot which shows the change in quenching efficiency relative to quencher concentration [152]. This approach was used for example to examine fesselin, which was shown to be a natively unfolded protein by fluorescence quenching, among other methods [153].

Another fluorescence spectroscopy based method especially useful for studying the presence of hydrophobic cores in an IDP, is the measurement of the excitation spectra after binding of 1-anilino-8-naphthalenesulfonic acid (ANS). When ANS binds to a hydrophobic pocket of a loosely folded protein or a partly denatured protein it causes a blue shift in the emission maximum and an increase in the intensity. The strength of this method is that it enables the separation of a fully extended random-coil like IDP from a molten globule type IDP since the molten globule will have residual hydrophobic core-like structures that will be exposed upon denaturation [135, 154-156].

3.4.4 Sodium dodecyl sulfate polyacrylamide gel electrophoresis, size exclusion chromatography, and limited proteolysis

Intrinsically disordered proteins have an unusual mobility on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) [92, 157]. The mobility is retarded so that IDPs usually appear to be at least 10% larger than expected based on their amino acid sequence. This phenomenon is thought to occur because of the unusual amino acid composition which reduces the SDS binding to the protein surface and, therefore, the mobility in the gel. A similar type of behavior is observed in size exclusion chromatography experiments, but for a different reason. In this case, the less compact, disordered, flexible protein has a larger hydrodynamic radius which results in the protein appearing larger than expected [135].

Limited proteolytic digestion is based on SDS-PAGE separation of a proteolytically fragmented protein [158, 159]. Proteinases with broad substrate specificity such as trypsin, proteinase K, and pepsin are commonly used [159]. The protein of interest

is first digested under limited conditions which can be achieved by restricting the time of the digestion, the amount of proteinase or by lowering the reaction temperature. After fragment separation the N-terminus or the masses of the fragments can be determined. Localization of the fragments enables the estimation of flexible or exposed regions and protruding loops. The use of this method can be extended to study conformational changes and different structural states of the protein by varying the conditions or by adding interaction partners. For example, the formation of a certain structural intermediate can be induced by changing the pH conditions as in the case of apomyoglobin [160]. It has an acidic molten-globule state at pH 4.2 and a native state at pH 8.0. Adjustment of the pH enabled the determination of which helices were disordered and which were stable [161].

3.4.5 Bioinformatics

Bioinformatics is perhaps the most convenient way to test whether a protein of interest could have intrinsically disordered features. Basic computer skills are enough to use most of the prediction softwares and only the amino acid sequence is required for input. Several types of software are available and many of those generate high levels of reliability (~85%) [97, 137]. PONDR® (Predictor of Natural Disordered Regions) is among the most widely used [106, 162, 163]. It is a collection of disorder predictors based on neural networks trained on specific sets of ordered and disordered proteins. The length of the disordered region can affect the prediction accuracy, and, because of this, the PONDR predictor package was recently updated with novel length dependent predictor algorithms [164, 165]. Initially, predictors use the amino acid sequence to calculate certain values such as hydropathy and charge for each residue in windows of, for example, 21 amino acids. These values are fed to the neural network which returns a value for each residue. If the value exceeds a certain threshold, the amino acid is considered disordered. The neural networks are trained with data from solved x-ray structures. However, instead of using the stable regions of these structures, disorder predictors, such as DISOPRED and Disembl [166] use the coordinates missing from electron density maps of x-ray structures. Some of the most commonly used predictors are briefly described and compared in Fig. 4.

To back up the prediction of the disordered regions, other bioinformatic tools can be used. Disordered proteins usually have typical charge profiles and this has been utilized by Uversky *et. al.* to provide a charge-hydropathy plot where proteins are positioned according to their mean net charge and hydropathy [95]. This approach does not give a per residue estimation of the disorder, but can be used as a more general classification method for IDPs. Another simpler approach is to count each amino acid and divide them into groups of disorder and order promoting residues. This can be automated using a composition profiler tool [167]. Certain amino acids, such as lysine, arginine, serine, proline, and glutamic acid, are potent generators of disorder, and enrichment of these in the amino acid sequence can be considered as an indication of intrinsic disorder [105, 106]. Although bioinformatics is an easy and quite reliable tool to identify and even localize intrinsic disorder in proteins, it can not replace wet lab experimentation which should always be used to verify initial *in silico* results [136].

Eventually, the combination of, preferably, several wet lab methods and different types of prediction software could lead for example to the localization of the disordered region, revealing an interaction surface or hinge between two domains. One may also find that a domain or an entire protein has an ensemble of several transient structures providing a landscape of structural intermediates. All this is indispensible information when deciphering the functional network of a protein.

AIMS OF THE STUDY

The aim of this study was to assess the key biochemical properties of PVA VPg. These properties were put into a structural context and the with a goal to evaluate the structure – function relationships. Since VPg turned out to be an intrinsically disordered protein, the aim for further structural studies was to identify structure stabilizing interactions and to characterize the nature of the stabilization.

MATERIALS AND METHODS

A detailed descriptions of the methods are in the original publications, as listed below.

Materials and Methods in Original Publications

Method	Original publication
Agroinfiltration	II
Bioinformatic analysis	I and II
CD spectroscopy	I and III
Electron microscopy	III
Fluorescence spectroscopy	I
Limited trypsin proteolysis	I and III
Luciferase assay	II
NMR	I
NTP binding assay	II
PIP strip binding assay	III
Plant expression vector cloning	II
Polyacrylamide gel electrophoresis (PAGE)	I
Recombinant protein production	I
RNase assay	II
Translation inhibition assay	II
Uridylylation assay	II
Vesicle preparation	III

RESULTS AND DISCUSSION

1. Biochemical properties of PVA VPg

This study used His–tagged recombinant VPg in most experiments. The VPg protein was produced in the E. coli M15 strain and purified under denaturing conditions. The folding of re-natured protein was compared to that of VPg purified under native conditions to verify the integrity of the structure (Fig. 1D in **I**). Since the folding seemed to be similar and independent of the purification method, and the denaturing purification led to higher recovery of purified protein, the denaturing purification protocol was used throughout these studies. The reversibility of the folding of many IDPs is, in fact, an advantage for recombinant protein purification since impurities may be irreversibly aggregated and removed. Typically, two major bands corresponding to the sizes of the VPg monomer and dimer were detected in all purifications under denaturing or native conditions (Fig. 1A, B and C in **I**). Purified samples lacked any major impurities but higher oligomers were occasionally detected (Fig. 1B in **I**). Dimers were also detected from plant samples where VPg was expressed using the 35S promotor (Fig. 5A in **II**). Dimerization complicated some analyses as did the high isoelectric point (~8.9) of the protein, as well as the pronounced aggregation near physiological pH. The high pI was a consequence of enrichment of positively charged Lys and Arg residues in the N-terminus. These properties are indicative of a need to certain biochemical environment and to regulation of the structural properties of the protein in a biological environment. In this sense, it could be argued that the in vitro setup of the study was highly artificial, but, in fact, this approach gave an easily manageable starting point for structural and biochemical studies of a dynamic protein. Purity or yield was not an issue and recombinant VPg behaved well in almost any buffer solution with a pH below 6.5. The fact that most of the biophysical methodology would have been impossible to conduct in vivo also made the in vitro approach an unavoidable compromise.

1.1 PVA VPg in NTP–binding and uridylylation

The amino acids 38-44 (AYTKKGK) are important for NTP–binding and uridylylation of VPg [16]. In this study, we inspected this region further and mutated the three lysines to alanines. The goal was to study the biochemical and structural properties of the binding and uridylylation processes in detail. Chemical cross-linking with sodium cyanoborohydrate was used to study the binding, as described in **II** and in the references therein [16,168]. The binding of UTP and the uridylylation efficiency was measured after SDS-PAGE separation as radioactivity from [α–^{32}P]UTP incorporated into VPg. The fact that the binding assay was based on lysine specific chemical cross-linking of the nucleotide means that it was not possible to determine whether VPg has an actual affinity for UTP or to calculate the Kd of the reaction. By definition, the nucleotide binding activity of a protein is a selective and non-covalent interaction (gene ontology definition GO:0000166 for nucleotide binding). Therefore, the term binding assay does not describe this setup very well. However, by this approach we can see whether the mutated lysines are available for chemical crosslinking, and, perhaps, form conclusions regarding the effect of the surface charge or local conformation of this lysine rich region of VPg to the approaching UTP.

PVA VPg is known to bind RNA non-specifically [169]. The mutated N-terminal region of the protein has a high positive charge which makes it a likely candidate for

RNA or nucleotide binding. We found that the cumulative alanine mutations gradually decreased the in vitro NTP–binding and uridylylation (Fig. 3 in **II**). This result supports the hypothesis that the lysines take part in these two activities; however, this did not enable us to determine whether VPg or NIb possesses an affinity for UTP. The uridylylated tyrosine is most probably Tyr63, as shown for the corresponding tyrosines in two other potyviruses, the Tobacco vein mottling virus and the Pepper vein banding virus [17, 19]. The structures of these VPgs are not known, but the poliovirus VPg structure and the model for uridylylation based on it suggests a coordinating effect for the lysine rich region rather than binding [21, 22]. Based on the results presented in Fig. 2 of **II** and the model of poliovirus uridylylation it appears that the flexibility of the disordered and charged NTP-binding region directs the approaching UTP to the proximity of the Tyr63 together with the coordination aid from manganese. In fact, the corresponding lysines in poliovirus VPg were recently shown to affect uridylylation in a similar way and are also suggested to be in direct contact with the poliovirus polymerase [170]. Therefore, in light of current knowledge, the NTP–binding activity associated with the uridylylation reaction may be a feature of the polymerase and not that of VPg, whose lysine rich region would serve in NTP-coordination. The structural assessment of such a small region is difficult, but a bioinformatic analysis suggests that this loss of activity could be related to a reduction in the flexibility of the region (Fig. 6 in **II**). Such a change would explain the lowered availability of the region for UTP cross-linking as well as the weakened uridylylation reaction as the coordination provided by the region would be disturbed. These changes could prevent the formation of the replication complex and explain the loss of infectivity (Fig. 2 in **II**).

1.2 PVA VPg in translation inhibition

The interaction of VPg with eukaryotic initiation factors is strong evidence for its participation in translation related processes [42, 55]. The in vitro translation inhibition activity of potyviral VPg has been shown in several studies [45, 54, 56], clearly demonstrating a role for VPg in translation. We examined this role in more detail and compared it to other activities by setting up luciferase activity based systems for the WT and the above mentioned lysine mutant VPgs. These systems were used for studying infectivity, as well as in vitro and in vivo translation inhibition. The constructs used are described in the Materials and Methods section of **II** in detail.

The in vitro translation inhibition activity of PVA VPg on capped mRNAs was gradually weakened when lysines in the NTP–binding region were mutated (Fig. 4B in **II**). In contrast, no clear dependence on lysines was detected when non-capped mRNA was used, suggesting that there could be a cap–dependent mechanism for the translation inhibition activity. The in vivo translation inhibition activity was tested using two luciferase reporter genes. VPg was expressed in N. benthamiana 24 h prior to infiltration of the Fluc and Rluc constructs. VPg expression resulted in the translation inhibition of both reporters (Fig. 5B and C in **II**). Interestingly, the lysine mutations did not affect the translation inhibition in vivo. This result may indicate that the translation inhibition activity of VPg is not controlled by the NTP–binding region in vivo or that this activity is specific to non–capped mRNAs. Although the in vivo assay showed inhibition of translation, it is difficult to draw conclusions regarding the exact mechanism based on this assay. For example, the role of the eIF4E interaction or the silencing suppression activity could be related to translation inhibition, but these connections remain to be studied.

2. Structural properties of PVA VPg

A far–UV CD spectrum was the first evidence of intrinsic disorder in the PVA VPg (Fig. 1D in **I**). The spectrum exhibited the typical features of IDPs introduced earlier. Based on the ellipticity of approximately -4000 deg/cm^2/dmol^{-1} at 222 nm and approximately -6000 deg/cm^2/dmol^{-1} at 200 nm (Fig. 1D in **I** and Fig. 2), VPg behaves like a molten globule–like IDP based on the classification presented by Uversky [99]. However, the more general term ´intrinsically disordered´ is used throughout this work, as in the title of the original publication of the structural properties of PVA VPg (**I**). The structural environment was further examined using CD spectroscopic analysis of the denaturation profiles. Heat denaturation revealed a relatively low melting temperature of ~42 °C and a reversibility of the folding (Fig. 2B and C in **I**). The deepening of the minima and the blueshift from 205 nm to 200 nm indicates a collapse of residual secondary structure to random coil upon heating (Fig. 2 and Fig. 2B in **I**). When the secondary structure proportions of these two forms were calculated, the collapsed structures appear to contain both sheets and helices (Table 1 in **I**). However, in the folded form of VPg the α-helix content was expected to be ~20% higher on the basis of the secondary structure predictions presented in Fig. 6 in **I**. Together with the trifluoroacetic acid (TFE) induced increase in α–helix content presented by Grzela et al. [6], these results suggest that stabilization of the disorder in potyviral VPgs could lead to an increase in α–helix content.

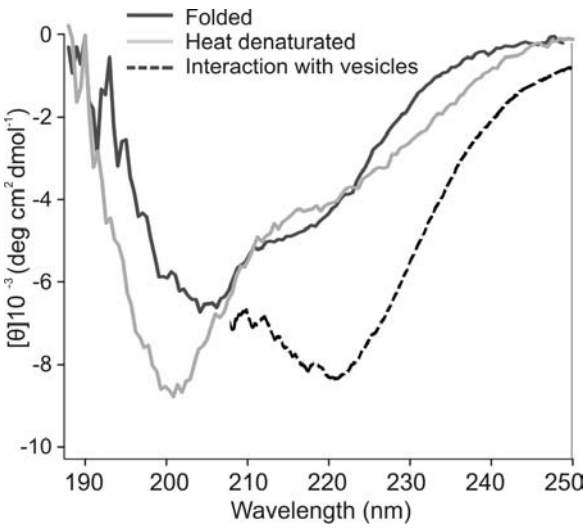


Figure 2. *Comparison of CD spectra of folded, heat denatured, and vesicle interacting VPg, originally published in* ***I*** *and* ***III****. The comparison shows the three different folds of VPg where the denatured form has the lowest amount of secondary structure, as indicated by the strong minimum close to 200 nm, and the vesicle stabilized form shows an increase in α–helical content, as indicated by the minimum shift to 222 nm.*

Studies using chemical denaturation and near-UV spectral analysis revealed that the tertiary structure, possibly including dimers, was disrupted at relatively low GdmHCl concentrations. This result indicates that a loosely folded structural intermediate state, devoid of tertiary structure but with pronounced secondary structure, may accumulate in these conditions (Figs. 2 and 3 in **I**). Evidence for disruption of a hydrophobic core domain was provided by NMR and fluorescence spectroscopy studies (Figs 4 and 5 in **I**). Binding of ANS was affected at low GdmHCl concentrations as was the tertiary structure observed in the near–UV region by CD spectroscopy. The hydrophobic core domain could therefore be part of a dimer or higher oligomer containing tertiary structure. Still, the expectation of structural, and more precisely α-helical, stabilization was evident throughout the studies of the disordered state of VPg.

2.1 The proportion of disorder in VPgs

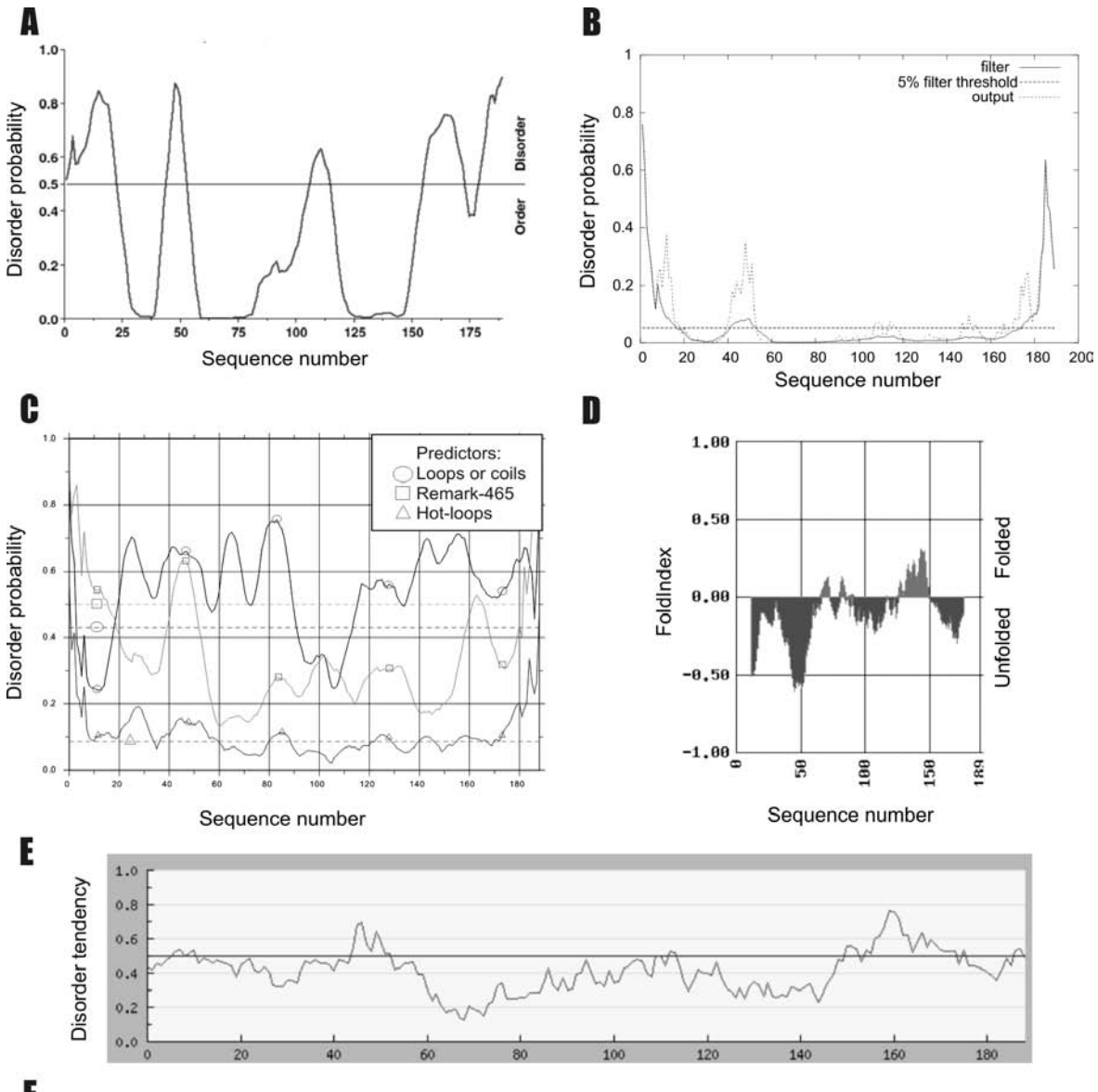
Several prediction and sequence analysis tools were used in this work. The PONDR® predictor of disordered regions was used in **I** and **II** to estimate the positioning

and content of disordered regions. The bioinformatic analysis was extended to a few other programs to compare the localization and amount of predicted disorder and to get an update on this rapidly evolving branch of bioinformatics (Fig. 3A-E). Some differences in the localization of predicted disorder between different programs was noted, but all predicted some degree of disorder. The N-terminus was predicted to be disordered by all the programs, with the exception of IUPred, and most software predicted a few other disordered stretches in VPg (summarised in Fig. 3F). In conclusion it can be said that in addition to loosely folded regions, VPg probably has a significant amount of structured parts.

When the secondary structure content was calculated from the CD spectra data in **I**, approximately 48% of VPg was estimated to be random coil or unstructured. In comparison, the proportion of residues predicted to be disordered by PONDR was 40%. Taken together, a rough estimate is that half of the protein is disordered. This amount seems to be close to the average for disordered proteins, in general, when the proportion of disorder in potyvirus VPgs was compared to that of structured proteins, disordered proteins, and sobemovirus VPgs. (Table 2 and Fig. 4). The distribution of the predicted disorder proportion was high in the DisProt and PDB datasets as can be expected from randomly selected and heterologous sets of proteins. Both VPg datasets showed a higher proportion of disorder compared to the PDB dataset of proteins with solved structures. Sobemovirus VPgs seemed to possess an even higher amount of disorder than potyviral VPgs. From this analysis, together with the amino acid composition analysis conducted earlier (Fig. 7 in **I**), intrinsic disorder appears to be a common property for VPg proteins. The same conclusion was recently drawn in a similarly extended study with two other VPgs [7]. In addition to verifying the intrinsic disorder in Rice yellow mosaic virus (genus Sobemovirus) and Lettuce mosaic virus (genus Potyvirus) with wet lab experimentation, the bioinformatic part of that study was extended to 14 VPgs from other viruses, which were all predicted to contain disordered regions.

2.2 Localization of the disordered regions

The bioinformatic analysis provided some clues as to the localization of the disorder in VPg (Fig. 6 in **I** and Fig. 3). The N–terminal region seemed to be the strongest candidate. It has an unusually high number of positively charged residues. Limited trypsin digestion was used to examine the flexible and exposed regions of VPg. The N-terminus of VPg was readily cleaved which confirmed that this region is easily accessible and probably disordered. Two other trypsin cleavage sites also corresponded well with the short regions predicted to be unstructured by PONDR® (Fig. 6 A and B in **I**). Some interesting features of these regions have been noted in the literature. The first region (fragment 4 in Fig. 6. in **I**) is cleaved at Thr45 which is right next to the proposed NTP–binding site [16] and in the middle of the recently identified nuclear localization signal (NLS II) [59]. The second accessible cleavage site was Asn115, next to Met116, a residue which was shown to be essential for virulence probably through a movement related function [171]. This site is also within the proposed amphiphilic central helix which was suggested to take part in HcPro and eIF4E interactions [46]. This is a clear indication that accessible and loosely folded regions are important for VPg function.



F

Predictor	Disordered region predicted on VPg	Predictor based on*
PONDR	1-22, 44-52, 107-114, 155-172, 179-189	Local aa composition, felxibility, hydropathy, etc.
Disopred2	1-12, 177-189	Cascaded support vector machine classifiers trained on PSI-BLAST profiles
IUPred	45-53, 155-176	Energy resulting from interresidue interactions estimated from local aa composition
FoldInfex	1-65, 72-80, 91-115, 150-189	Charge/hydropathy analyzed locally using a sliding window
Disembl	Loops/Coils: 19-91, 115-189 Hot-loops: 1-60, 171-189 Remark-465 1-16, 42-51	Neural networks trained on X-ray structure data

***Figure 3**. Comparison of amounts of predicted disorder in PVA VPg and localization of the disordered regions. A) PONDR® [106, 163]. B) Disopred2 [212]. C) Disembl [213]. The Disembl predictor combines three separate definitions of disorder. The threshold line for each definition is given as a dotted line with corresponding symbol. D) FoldIndex [214]. E) Iupred [215]. F) Summary of the prediction results for PVA VPg and the basis of the predictor.*

Table 2. *Viral VPgs included in the datasets in Fig. 4.*

PotyVPg

Potato virus Y
Bean yellow mosaic virus
Blackberry virus Y
Chilli veinal mottle virus
Onion yellow dwarf virus
Wisteria vein mosaic virus
Daphne virus Y
Shallot yellow stripe virus
Thunberg fritillary virus
Narcissus degeneration virus
Lily mottle virus
Watermelon mosaic virus
Telosma mosaic virus
Peace lily mosaic virus
Banana bract mosaic virus
Turnip mosaic virus
Potato virus A
Plum pox virus
Pepper mottle virus
Zucchini yellow mosaic virus
Papaya ringspot virus
Sorghum mosaic virus
Peru tomato mosaic virus
Soybean mosaic virus
Lettuce mosaic virus
Tuberose mild mottle virus
Sweet potato feathery mottle virus
Leek yellow stripe virus
Cocksfoot streak virus
Papaya leaf-distortion mosaic potyvirus

SobemoVPg

Cocksfoot mottle virus
Turnip rosette virus
Sesbania mosaic virus
Ryegrass mottle virus
Rice yellow mottle virus
Lucerne transient streak virus
Southern bean mosaic virus
Subterranean clover mottle virus
Southern cowpea mosaic virus

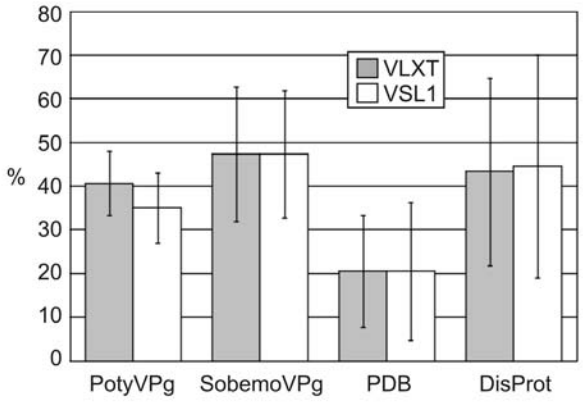


Figure 4. *Intrinsic disorder in potyviral and sobemoviral datasets compared to randomly collected PDB and DisProt datasets. Both random datasets included 30 proteins with chain length of 170 to 210 amino acids. The PONDR® software was used to estimate the total content of intrinsic disorder in potyvirus VPg, DisProt, and PDB datasets. In addition, a dataset of manually collected 9 sobemovirus VPgs was analyzed. Sequences of all proteins from all datasets were analyzed separately with VLXT and VSL1 algorithms using the PONDR® server. The number of residues predicted to be disordered was calculated and averaged to get an approximation of the proportion of disorder in a given dataset.*

2.3 Stabilization of intrinsic disorder

Commercial phosphatidylinositol strips or "PIP strips" (Echelon Biosciences) were used to test whether VPg interacts with membrane lipids. Despite the non-specific binding of antibody to several of the lipids on the strips, VPg showed specific binding to certain lipids and further experiments were based on these results (Fig 1. in **III**). CD spectroscopy was used to characterize the structure of VPg in the presence of synthetic lipid vesicles (Fig. 2 in **III**), and a change in structure was detected following interactions with the anionic phospholipids phosphatidylserine (DOPS) and phosphatidylglycerol (DOPG). The changes resembled those observed in PVY, *Rice yellow mosaic virus*, and *Lettuce mosaic virus* VPgs upon addition of TFE [6,7], which is known to promote α–helix formation [172]. Although the high lipid background prevented observations at 185 nm or 208 nm, which are wavelengths indicative of α–helices, the appearance of a pronounced minimum at 222 nm was evidence of α–helical stabilization (Fig. 2 in **III**). Three different conformational states of PVA VPg are presented in Fig. 2, where clear differences between the folded but partly disordered state, the heat denatured state, and the vesicle interaction stabilized state can be seen. Because CD spectroscopy can not determine the location of the elements undergoing stabilization, and because the change in structure needed to be independently verified, limited trypsin digestion was used to study the vesicle interacting form of VPg. The trypsin digestion patter of VPg in the presence of non-interacting phosphatidylcholine (DOPC) vesicles was identical to the pattern in the absence of vesicles, whereas, in the presence of DOPG and DOPS vesicles, an obvious change in the pattern was observed (Fig. 3 in **III**). The N-terminus of the VPg was predicted to be disordered, and so can it be, but since there is actually an increase in N-terminal fragmentation after vesicle association, the ~25 N-terminal amino acids do not appear to be protected in this interaction. The most notable change was the disappearance of the two most stable fragments cleaved at positions Thr45 and Asn115 in the control reactions. As mentioned above, these residues are in functionally intriguing positions and the disappearance of these in the lipid interacting samples suggests that they may be involved in an interaction that stabilizes α-helices. Based on the CD spectra and limited trypsination results, a change in VPg structure upon interaction with anionic vesicles was evident, but was there any changes in the vesicles? This was next examined by electron microscopy (EM).

2.4 Vesicle modifications caused by VPg – Evidence for a pore forming activity

EM was used to visualize the changes in vesicle appearance. No difference was seen in DOPC vesicles after VPg addition (Fig. 4B in **III**), whereas electron dense spots appeared on the vesicle surface in the DOPS and DOPG samples (Fig. 4D in **III** and Fig. 5). The spots in the images gave a strong impression of holes which seemed to lead to the disruption of the vesicle surface (Fig. 4E in **III** and fig. 5C and D). In addition to the samples described in the original publications, an alternative sample preparation method was tested to confirm that the observed changes were not sample preparation artifacts. These samples were prepared by cutting ultrathin sections of samples cast in resin and although the sample quality was poorer, similar observations were made (data not shown).

The appearance of the vesicles and the nature of the structural stabilization led to the hypothesis that VPg possesses a hole or a pore forming activity, as proposed in Fig. 6. The model partially follows the mechanisms proposed for the charged HIV-1 TAT peptide [173] and the carpet mechanism proposed by Shai for antimicrobial peptides [174-176]. In the carpet mechanism, the membrane lytic peptides form a localized carpet of

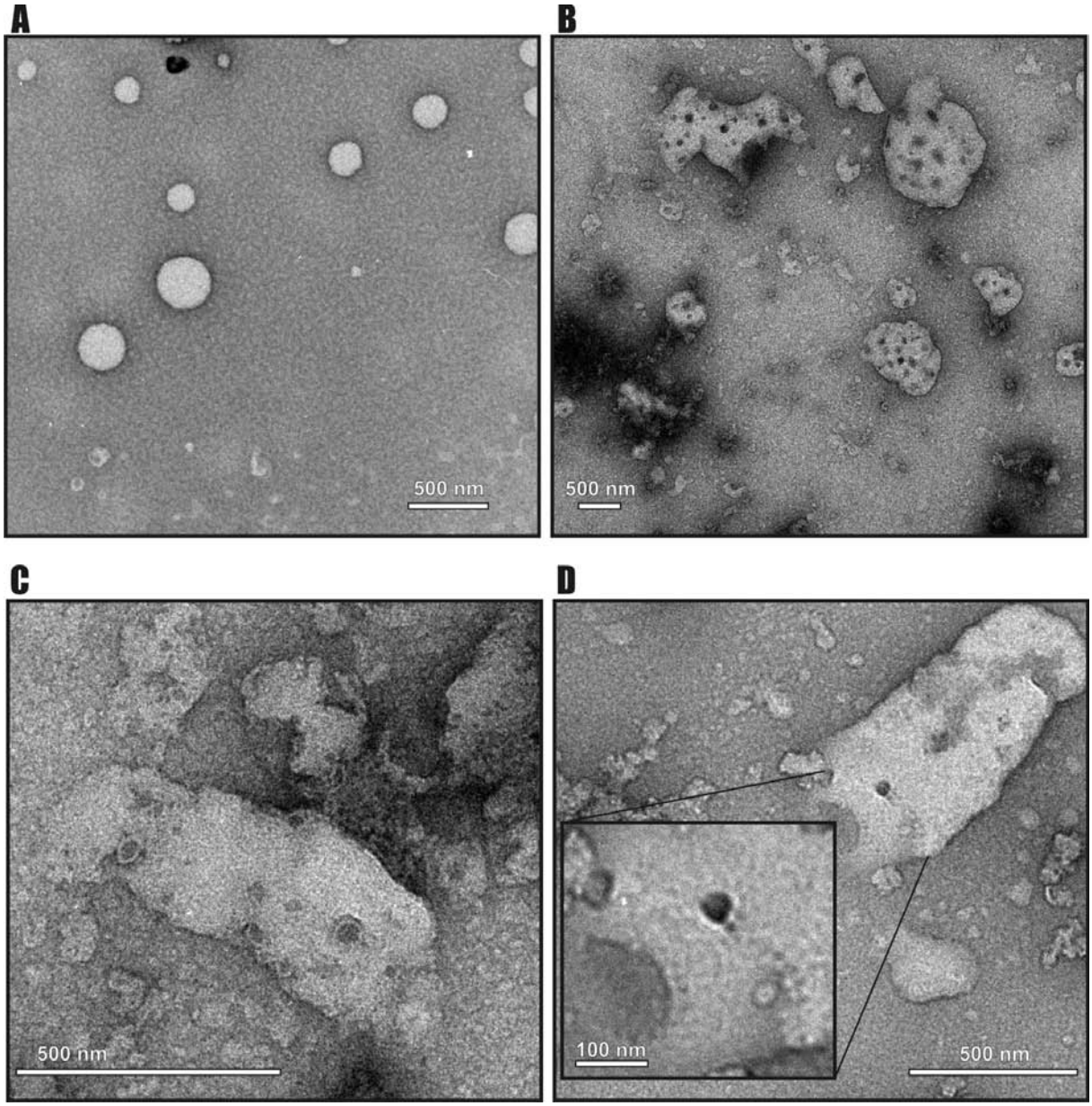


***Figure 5**. Electron microscopic images of VPg and DOPS vesicles. A) DOPS vesicles in the absence of VPg were symmetrical spheres with average size of ~100 nm (**I**) B) Addition of VPg and a 1 h incubation at room temperature led to formation of electron dense spots, an approximately five fold increase in size, and a disruption of the vesicle surface. C and D) With closer inspection the electron dense spots seemed to have pore–like properties. Additionally, micelle–like smaller vesicle structures appeared outside the swollen vesicles after VPg addition.*

concentrated protein which penetrates into the membrane bilayer and, eventually, leads to membrane lysis. There are several lines of evidence favoring this mechanism for PVA VPg. First, in the carpet mechanism, the initial interaction between the lipid and the α–helical peptide is driven by an electrostatic interaction, which is most probably the case between the negatively charged surface of the DOPS vesicles and positively charged patches in VPg. This mechanism does not require the amino acid chain to penetrate into the hydrophobic core of the membrane and would explain the strong binding of VPg to the vesicles in the absence of obvious transmembrane helices in the VPg sequence.

The carpet mechanism would also lead to dissociation of micelle–like particles from the membrane, as seen across the background in Fig. 5B and indicated with an ´m´ in Fig. 4D in **III**. Prolonged incubation of vesicles with VPg eventually led to disruption of the vesicle surface as seen in Fig. 5C and D. This outcome would also be expected for membrane lytic antimicrobial peptides in the carpet mechanism.

The model presented for VPg in Fig. 6 is also supported by the limited trypsin

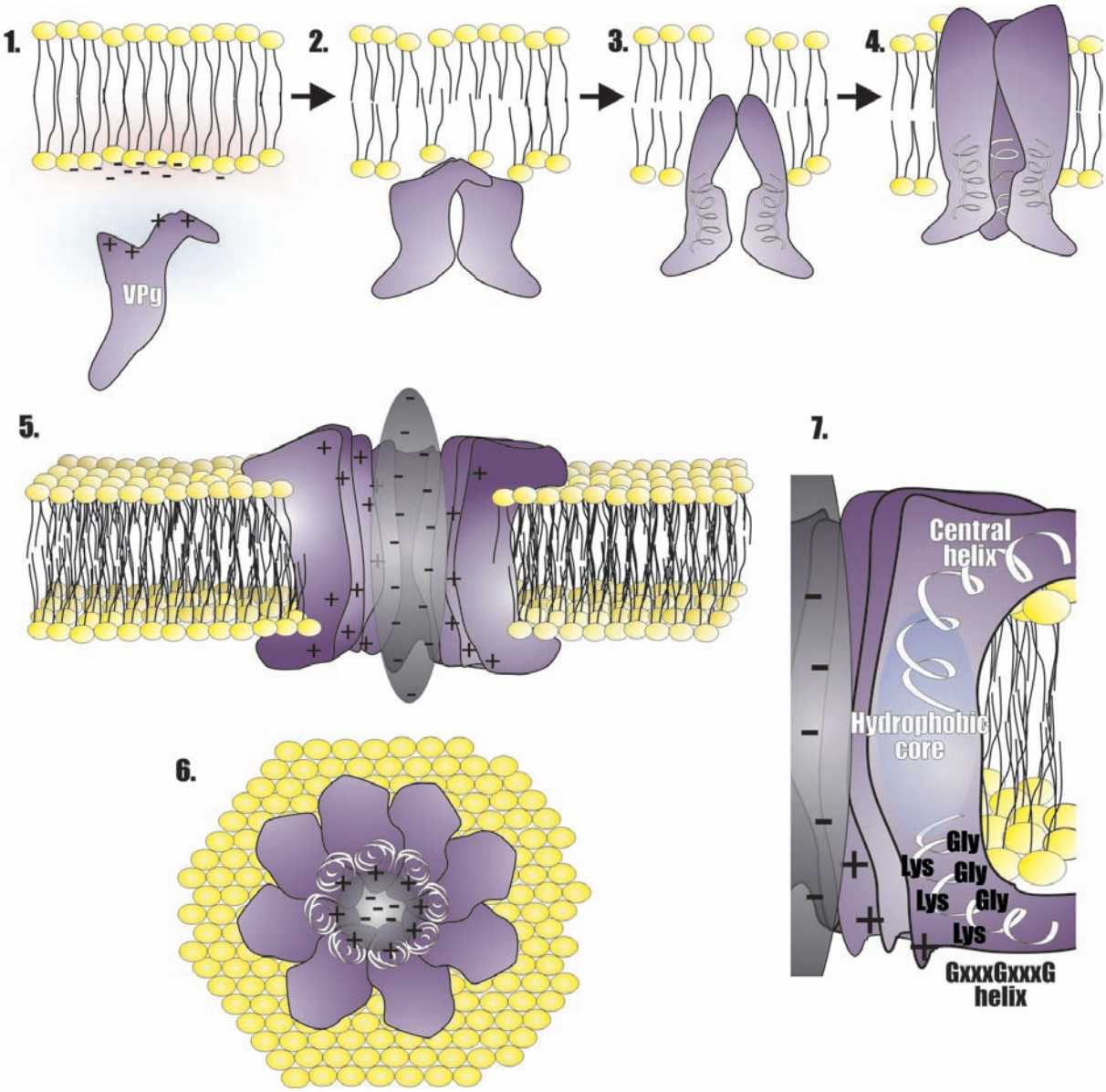


Figure 6. *Hypothetical model of membrane channel or pore formation by VPg. 1) VPg is attracted by opposite charges and binds to the surface of the anionic vesicle through electrostatic interactions. 2-4) The protein-vesicle interactions lead to α–helical stabilization of VPg secondary structure. Structural stabilization and the positive charges of the GxxxG motifs cause the VPg to penetrate the vesicle surface. This stage could resemble the carpet mechanism described for antimicrobial membrane lytic peptides [174]. 5-7) Structure stabilization and membrane penetration leads to formation of channel or pore–like structures with multiple VPg copies surrounding the pore. Amphipathicity of the stabilized helices would align the positive charges of the lysine and arginine side chains towards the center of the pore and the proposed hydrophobic core (**I** and Fig. 7) would cover the hydrophobic central region of the membrane bilayer. The result would be a pore with positive charges capable of transporting negatively charges molecules such as RNA across the bilayer.*

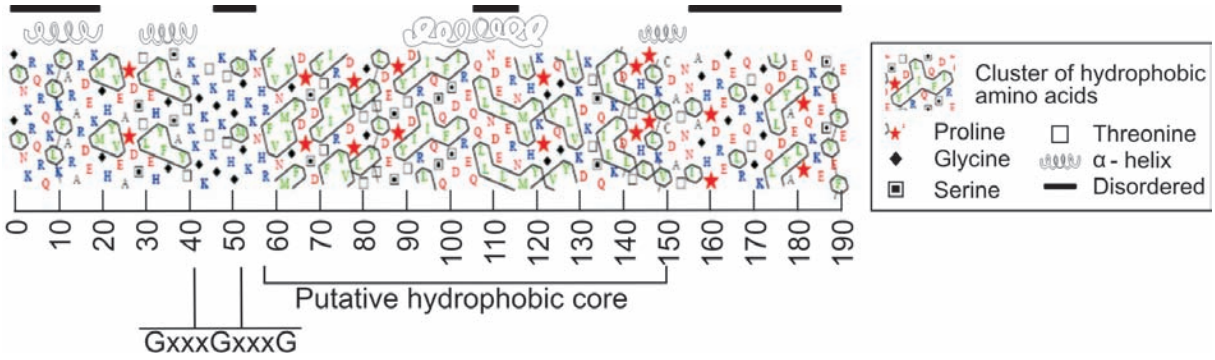


Figure 7. *Hydrophobic cluster analysis of PVA VPg showing the clustering of hydrophobic amino acids. Analysis method is based on a two-dimensional representation of helically mapped amino acid sequences. A detailed description of the method is presented in the original publication [216]. Predicted α–helices and disordered regions are indicated at the top of the figure. Secondary structure element and disorder positions are based on predictions made in* ***I****. The hydrophobic core domain proposed in* ***I*** *is suggested to be located in the central region.*

digestion results and the sequence properties of VPg. There are two positively charged patches in the N–terminal region of VPg and a large region with hydrophobic clusters approximately ranging from amino acids 60 to 150 (Fig. 7). Asn115, the trypsin cleavage site which was protected in the vesicle associated VPg, is in the middle of this region (Fig. 2 in **III**). This Asn residue is also in the middle of the predicted central helix which is involved in eIF4E and viral HcPro interactions (Fig. 6 in **I** and [46]). Additionally, this region is the most plausible choice for the hydrophobic core domain suggested in **I**. For all these suggestions to be valid and credible there should not be any conflicts or discrepancies. For example, is it possible that this region contains the hydrophobic core and the disordered but stabilized α–helix, and is involved in eIF4E and HcPro interactions? The scenario is made realistic by the disordered and flexible nature of VPg. It could allow, for example, the disordered central helix to be stabilized by the interaction with the membrane surface and associate with eIF4E or HcPro to start translation in complex with the other required components. This scenario is, of course, hypothetical, but theoretically plausible.

Another very interesting sequence feature favoring a pore forming activity via helical stabilization is the GxxxG motif in VPg. In fact, there is a conserved tandem repeat of this motif from Gly43 to Gly51. Substantial evidence shows that GxxxG motifs are tightly involved in membrane associated dimerization and pore formation by helix-helix interactions [177-179]. This type of domain is found in the hepatitis C virus where it plays a role in heterodimerization of the envelope proteins E1 and E2 [33] and in the SARS coronavirus spike protein where it enhances trimer formation. A bioinformatic study revealed that, based on 13 non-homologous membrane channel and transporter protein structures, glycine positioned at every fourth residue is the most preferred amino acid contributing to helix-helix contacts [180]. Another bioinformatic survey concluded that GxxxG and GxxxxxxG motifs are among the most prevalent contributors to transmembrane helices. A GxxxxxxG motif is located upstream from the first Gly of tandem GxxxG motif in the PVA VPg, forming a GxxxxxxGxxxGxxxG motif, in which the glycines are conserved among all potyviruses (data not shown). This motif is therefore the strongest candidate for a membrane helix forming region.

GxxxG mediated dimerization of the hepatitis C virus envelope proteins and trimerization of the SARS spike proteins

suggest that that PVA VPg multimerization could be associated with membrane interactions. Although VPg dimerization is well documented (e.g. Fig. 1 A to C in **I**) and higher oligomers are occasionally seen on SDS-PAGE gels, the functions of these multimers are unknown. The presence of VPg dimers was also verified in planta with antibody detection as shown in Fig. 5A in **II**. SDS-PAGE usually denaturates proteins and dissociates multimers but many membrane proteins are resistant to dissociation and maintain their multimer status [181]. For a pore or channel to form, several complexed proteins are required. A common mechanism of membrane insertion associated multimerization is through helix-helix bundles [182] which would fit well with the known biochemical properties of VPg.

2.5 Implications of membrane modifications for virus biology

Virus induced membrane modifications are most often vesicles with sizes that vary from 40 nm to a few micrometers. The origin of the lipids is usually the host ER, but other organelles, such as lysosomes or mitochondria, can also be used [29]. In some cases, these vesicles have been shown to have bottlenecks, channels or pores open to the cytoplasm [26,173]. A protein group designated viroporins consists of virus encoded proteins that are capable of expanding lipid bilayers and forming hydrophilic pores [183]. The purposes of these formations are diverse. Ions, RNA, nucleotides and proteins can be transported through these openings and provide trafficking of the vesicle contents.

When the vesicle interaction and membrane modifications of PVA VPg presented here are considered in relation to uridylylation and replication, the well known requirement for a membranous environment for the replication complex formation is inevitably considered [28, 29, 184]. This requirement is best known from the *Togaviridae, Coronaviridae, Picornaviridae*, and *Flaviviridae* viruses, but examples from *Potyviridae* also exist. TEV 6K2 protein was shown to be involved in replication complex formation from plant cell ER membranes [185, 186]. Subsequently, this process was pinpointed to ER exit sites (ERES) and shown to involve COPI and COPII coating machineries [187]. ER associated vesicle budding was induced during TuMV infection by the polyprotein intermediate 6K-VPg-Pro. In addition, this intermediate was shown to interact with eIF(iso)4E inside the vesicles [15]. In the same study, the VPg-Pro intermediate (which corresponds to NIa in PVA) was shown to localize predominantly in the nucleus. In that study, the polyprotein intermediate cleavage sites were mutated to prevent full processing. Furthermore, a study of TuMV revealed that its polymerase interacts with the host heat shock protein 70 as well as eukaryotic initiation factor 1A inside virus-induced vesicles [188,189]. Fully processed PVA VPg can be purified from virus infected plant membrane fractions showing that it also has a role *in vivo* in the membranous environment similar to the examples above [79]. Replication and translation are most certainly activities that require the support and organization provided by membrane association. Membrane bilayers also provide a mechanism by which to sort and concentrate molecules through selective pores, as suggested for VPg in Fig. 6.

The VPg – membrane interaction described in this study is probably closely related to the transmembrane anchoring properties of 6K2, which precedes VPg in the polyprotein [186]. Based on the amino acid sequence, it seems unlikely that VPg itself has hydrophobic stretches capable of forming transmembrane helices. A more plausible mechanism of VPg membrane interactions and modification would be penetration of the bilayer by the amphipathic helix as proposed in the mechanism shown in Fig 6. A similar type of amphipathic helix binding may occur with nsp1 from the *Semliki forest virus* [190]. Nsp1

is devoid of any obvious hydrophobic patches, but is tightly associated with membranes through an amphipathic helix. Similar to VPg, nsp1 function is associated with replication. Additionally, nsp1 has a virus RNA capping activity and requires anionic phospholipids for its activity [30,191]. Another viral protein forming amphipathic helices on membrane surfaces is the tomato ringspot virus NTB-VPg protein, which has both a transmembrane anchor and an amphipathic helix in the same protein [36]. By analogy, it is possible that the PVA polyprotein intermediate is anchored tightly through 6K2 to the membrane until the proteolytic digestion of the polyprotein is complete. The role of the PVA polyprotein intermediates remains poorly understood, but is most probably an important part of the membrane associated functions of VPg.

CONCLUDING REMARKS

Some of the key VPg related virus infection cycle events are not well understood (Fig. 1). It is not clear, for example, whether replication occurs in the same environment as translation. Could it be that all these events take place in the same virus induced compartment? The membranous environment formed in the early stage of virus entry could provide support for several downstream activities. For example, before replication, an initial round of translation to produce the viral components for associated activities must take place. The initial translation has to occur on the ER associated translation machinery. Transcription, replication, and translation all produce molecules that have to be transported to the required locations. This includes transport through the membrane bilayer using pores or disruption of the membrane structures and the subsequent leakage of the contents. Based on the results presented here, it seems that in the midst of these activities there is a multifunctional and structurally flexible VPg which could be called the hub of PVA proteins.

ACKNOWLEDGEMENTS

This work was carried out at the department of Applied Biology and at the department of Applied Chemistry and Microbiology at the University of Helsinki during years 2005-2009.

I would like to thank Kristiina Mäkinen for the support and guidance through the PhD studies. Your understanding of biochemistry and virology has been an indispensable help for me. This combined with your kindness, and flexibility makes you a supervisor and a group leader that is easy to work with and difficult to leave. I would also like to thank my other past and current plant virus research group co-workers Katri Eskelin, Konstantin Ivanon, Anders Hafrén, Taina Suntio, Rasa Gabrenaite-Verkhovskaya, Eva Wahlström, Pietri Puustinen and Jane Besong. I would especially like to thank Kostya, who, in addition of showing me how experiments should be done, showed me what interesting and polite persons scientists can be. I would like to thank especially Katri for all the help, support and tips for cloning… and also for keeping up the order in the lab. Special thanks also for Anders for all the interesting, funny and intense conversations. You have been a great person to work and travel with and I have been really lucky to be able to share my PhD study time with you. I would also like to thank my former class-mates and colleaques from the University of Kuopio, Brahim Semane, Timo Nykopp, Mikko Sairanen and Miika Heinonen for helping me to adapt to life at the University.

I would like to acknowledge my other collaborators in science, who have showed me the significance and importance of collaboration. Professors Vladimir Uversky, Keith A. Dunker, Daniel Otzen and Doctors Perttu Permi, Nisse Kalkkinen and Peter Christensen. Completion of this thesis would not have been possible without your help.

I'm grateful for the financial support provided by the Academy of Finland and Viikki Graduate School in Biosciences. I thank VGSB coordinators Drs. Eeva Sievi and Sandra Falck for being a great help along the studies. I also wish to acknowledge the follow-up group members Prof. Timo Hyypiä and Dr. Roman Tuma for helping to choose which way to progress with the research. I also wish to thank Professors Paavo Kinnunen and Erkki Truve for the scientific review of this thesis.

Extra special thanks to my good old friends. Miika H., Roope I., Kalle J. (and Kuakeli for desktop publishing of this book), Anna-Kristiina K., Anssi K., Jussi K., Ville R., Jussi T., Kustaa V., and Aaro V. Your friendship is officially acknowledged and appreciated. Your help has been crucial in taking my mind out of my work.

Regardless of the overuse of special thanks I want to give extra colossal thanks to my family. I want to thank my parents Markku and Paula for telling me that education is important. I'm truly happy that I took your word and studied. You never put any pressure on my studies and that's clearly one of the reasons why I still like to study. Concerning the studies, I also have to thank you for buying us those computers. Knowing how to use them has really made my life easier... at least after we stopped fighting for the same one with my brothers. As we are past the fighting these days, I also want to thank my brothers Mikko, Panu and Tommi. I thank you as a standard procedure for being my brothers, but gladly and sincerely because you are also my good friends and a great company! I wish all the best for you and your astonishingly simultaneously growing families. Finally, I would like to thank Anni J. for support, tolerance, patience, caring, and sharing your life with me for the past few years.

REFERENCES

[1] Gibbs, AJ, Ohshima, K, Phillips, MJ, and Gibbs, MJ. (2008). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS One* 3, e2523.

[2] Hull, R. (2009). *Comparative Plant Virology*, 2nd edition, Academic Press.

[3] Shukla, DD, Ward, CW, and Brunt, AA. (1994) *The Potyviridae*, C.A.B. International, Wallingford, UK.

[4] Chung, BY, Miller, WA, Atkins, JF, and Firth, AE. (2008). An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5897-902.

[5] Satheshkumar, PS, Gayathri, P, Prasad, K, and Savithri, HS. (2005). "Natively unfolded" VPg is essential for Sesbania mosaic virus serine protease activity. *J. Biol. Chem.* 280, 30291-300.

[6] Grzela, R, Szolajska, E, Ebel, C, Madern, D, Favier, A, Wojtal, I, Zagorski, W, and Chroboczek, J. (2008). Virulence factor of potato virus Y, genome-attached terminal protein VPg, is a highly disordered protein. *J. Biol. Chem.* 283, 213-21.

[7] Hébrard, E, Bessin, Y, Michon, T, Longhi, S, Uversky, VN, Delalande, F, Van Dorsselaer, A, Romero, P, Walter, J, Declerk, N, and Fargette, D. (2009). Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. *Virol. J.* 6, 23.

[8] Léonard, S, Viel, C, Beauchemin, C, Daigneault, N, Fortin, MG, and Laliberté, JF. (2004). Interaction of VPg-Pro of turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta. *J. Gen. Virol.* 85, 1055-63.

[9] Murphy, JF, Rhoads, RE, Hunt, AG, and Shaw, JG. (1990). The VPg of tobacco etch virus RNA is the 49-kDa proteinase or the N-terminal 24-kDa part of the proteinase. *Virology* 178, 285-8.

[10] Oruetxebarria, I, Guo, D, Merits, A, Mäkinen, K, Saarma, M, and Valkonen, JP. (2001). Identification of the genome-linked protein in virions of Potato virus A, with comparison to other members in genus Potyvirus. *Virus Res.* 73, 103-12.

[11] Merits, A, Rajamäki, ML, Lindholm, P, Runeberg-Roos, P, Kekarainen, T, Puustinen, P, Mäkelainen, K, Valkonen, JP, and Saarma, M. (2002). Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *J. Gen. Virol.* 83, 1211-21.

[12] Guo, D, Rajamäki, ML, Saarma, M, and Valkonen, JP. (2001). Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system. *J. Gen. Virol.* 82, 935-9.

[13] Lin, L, Shi, Y, Luo, Z, Lu, Y, Zheng, H, Yan, F, Chen, J, Chen, J, Adams, MJ, and Wu, Y. (2009). Protein-protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Res.* 142, 36-40.

[14] Yambao, ML, Masuta, C, Nakahara, K, and Uyeda, I. (2003). The central and C-terminal domains of VPg of Clover yellow vein virus are important for VPg-HCPro and VPg-VPg interactions. *J. Gen. Virol.* 84, 2861-9.

[15] Beauchemin, C, Boutet, N, and Laliberté, JF. (2007). Visualization of the interaction between the precursors of VPg, the viral protein linked to the genome of turnip mosaic virus, and the translation eukaryotic initiation factor iso 4E in Planta. *J. Virol.* 81, 775-82.

[16] Puustinen, P and Mäkinen, K. (2004). Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. *J. Biol. Chem.* 279, 38103-10.

[17] Anindya, R, Chittori, S, and Savithri, HS. (2005). Tyrosine 66 of Pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. *Virology* 336, 154-62.

[18] Murphy, JF, Klein, PG, Hunt, AG, and Shaw, JG. (1996). Replacement of the tyrosine residue that links a potyviral VPg to the viral RNA is lethal. *Virology* 220, 535-8.

[19] Murphy, JF, Rychlik, W, Rhoads, RE, Hunt, AG, and Shaw, JG. (1991). A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. *J. Virol.* 65, 511-3.

[20] Schaad, MC, Haldeman-Cahill, R, Cronin, S, and Carrington, JC. (1996). Analysis of the VPg-proteinase (NIa) encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *J. Virol.* 70, 7039-48.

[21] Schein, CH, Oezguen, N, Volk, DE, Garimella, R, Paul, A, and Braun, W. (2006). NMR structure of the viral peptide linked to the genome (VPg) of poliovirus. *Peptides* 27, 1676-84.

[22] Schein, CH, Volk, DE, Oezguen, N, and Paul, A. (2006). Novel, structure-based mechanism for uridylylation of the genome-linked peptide (VPg) of picornaviruses. *Proteins* 63, 719-26.

[23] Wang, X, Ullah, Z, and Grumet, R. (2000). Interaction between zucchini yellow mosaic potyvirus RNA-dependent RNA polymerase and host poly-(A) binding protein. *Virology* 275, 433-43.

[24] Dufresne, PJ, Ubalijoro, E, Fortin, MG, and Laliberté, JF. (2008). Arabidopsis thaliana class II poly(A)-binding proteins are required for efficient multiplication of turnip mosaic virus. *J. Gen. Virol.* 89, 2339-48.

[25] Flanegan, JB and Baltimore, D. (1977). Poliovirus-specific primer-dependent RNA polymerase able to copy poly(A). *Proc. Natl. Acad. Sci. U. S. A.* 74, 3677-80.

[26] Kopek, BG, Perkins, G, Miller, DJ, Ellisman, MH, and Ahlquist, P. (2007). Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. *PLoS Biol.* 5, e220.

[27] Takegami, T, Kuhn, RJ, Anderson, CW, and Wimmer, E. (1983). Membrane-dependent uridylylation of the genome-linked protein VPg of poliovirus. *Proc. Natl. Acad. Sci. U. S. A.* 80, 7447-51.

[28] Denison, MR. (2008). Seeking membranes: positive-strand RNA virus replication complexes. *PLoS Biol.* 6, e270.

[29] Miller, S and Krijnse-Locker, J. (2008). Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* 6, 363-74.

[30] Salonen, A, Ahola, T, and Kääriäinen, L. (2005). Viral RNA replication in association with cellular membranes. *Curr. Top. Microbiol. Immunol.* 285, 139-73.

[31] Suhy, DA, Giddings, TH,Jr, and Kirkegaard, K. (2000). Remodeling the endoplasmic reticulum by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-induced vesicles. *J. Virol.* 74, 8953-65.

[32] Lama, J and Carrasco, L. (1996). Screening for membrane-permeabilizing mutants of the poliovirus protein 3AB. *J. Gen. Virol.* 77, 2109-19.

[33] Op De Beeck, A, Montserret, R, Duvet, S, Cocquerel, L, Cacan, R, Barberot, B, Le Maire, M, Penin, F, and Dubuisson, J. (2000). The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J. Biol. Chem.* 275, 31428-37.

[34] Kujala, P, Ikäheimonen, A, Ehsani, N, Vihinen, H, Auvinen, P, and Kääriäinen, L. (2001). Biogenesis of the Semliki Forest virus RNA replication complex. *J. Virol.* 75, 3873-84.

[35] Voisset, C and Dubuisson, J. (2004). Functional hepatitis C virus envelope glycoproteins. *Biol. Cell.* 96, 413-20.

[36] Zhang, SC, Zhang, G, Yang, L, Chisholm, J, and Sanfaçon, H. (2005). Evidence that insertion of Tomato ringspot nepovirus NTB-VPg protein in endoplasmic reticulum membranes is directed by two domains: a C-terminal transmembrane helix and an N-terminal amphipathic helix. *J. Virol.* 79, 11752-65.

[37] Pogany, J, Stork, J, Li, Z, and Nagy, PD. (2008). In vitro assembly of the Tomato bushy stunt virus replicase requires the host Heat shock protein 70. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19956-61.

[38] dos Reis Figueira, A, Golem, S, Goregaoker, SP, and Culver, JN. (2002). A nuclear localization signal and a membrane association domain contribute to the cellular localization of the Tobacco mosaic virus 126-kDa replicase protein. *Virology* 301, 81-9.

[39] Pestova, TV, Kolupaeva, VG, Lomakin, IB, Pilipenko, EV, Shatsky, IN, Agol, VI, and Hellen, CU. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7029-36.

[40] Wittmann, S, Chatel, H, Fortin, MG, and Laliberté, JF. (1997). Interaction of the viral protein genome linked of turnip mosaic potyvirus with the translational eukaryotic initiation factor (iso) 4E of Arabidopsis thaliana using the yeast two-hybrid system. *Virology* 234, 84-92.

[41] Schaad, MC, Anderberg, RJ, and Carrington, JC. (2000). Strain-specific interaction of the tobacco etch virus NIa protein with the translation initiation factor eIF4E in the yeast two-hybrid system. *Virology* 273, 300-6.

[42] Khan, MA, Miyoshi, H, Ray, S, Natsuaki, T, Suehiro, N, and Goss, DJ. (2006). Interaction of genome-linked protein (VPg) of turnip mosaic virus with wheat germ translation initiation factors eIFiso4E and eIFiso4F. *J. Biol. Chem.* 281, 28002-10.

[43] Le, H, Browning, KS, and Gallie, DR. (2000). The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B. *J. Biol. Chem.* 275, 17452-62.

[44] Duprat, A, Caranta, C, Revers, F, Menand, B, Browning, KS, and Robaglia, C. (2002). The Arabidopsis eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.* 32, 927-34.

[45] Grzela, R, Strokovska, L, Andrieu, JP, Dublet, B, Zagorski, W, and Chroboczek, J. (2006). Potyvirus terminal protein VPg, effector of host eukaryotic initiation factor eIF4E. *Biochimie* 88, 887-96.

[46] Roudet-Tavert, G, Michon, T, Walter, J, Delaunay, T, Redondo, E, and Le Gall, O. (2007). Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPro. *J. Gen. Virol.* 88, 1029-33.

[47] Michon, T, Estevez, Y, Walter, J, German-Retana, S, and Le Gall, O. (2006). The potyviral virus genome-linked protein VPg forms a ternary complex with the eukaryotic initiation factors eIF4E and eIF4G and reduces eIF4E affinity for a mRNA cap analogue. *FEBS J.* 273, 1312-22.

[48] Alberts, B, Johnson, A, Lewis, J, Raff, M, Rberts, K, and Walter, P. (2008) *Molecular biology of the cell*, 5th edition, Garland Science, New York, USA.

[49] Goodfellow, IG and Roberts, LO. (2008). Eukaryotic initiation factor 4E. *Int. J. Biochem. Cell Biol.* 40, 2675-80.

[50] Dreher, TW and Miller, WA. (2006). Translational control in positive strand RNA plant viruses. *Virology* 344, 185-97.

[51] Kneller, EL, Rakotondrafara, AM, and Miller, WA. (2006). Cap-independent translation of plant viral RNAs. *Virus Res.* 119, 63-75.

[52] Basso, J, Dallaire, P, Charest, PJ, Devantier, Y, and Laliberté, JF. (1994). Evidence for an internal ribosome entry site within the 5' non-translated region of turnip mosaic potyvirus RNA. *J. Gen. Virol.* 75 (Pt 11), 3157-65.

[53] Niepel, M and Gallie, DR. (1999). Identification and characterization of the functional elements within the tobacco etch virus 5' leader required for cap-independent translation. *J. Virol.* 73, 9080-8.

[54] Cotton, S, Dufresne, PJ, Thivierge, K, Ide, C, and Fortin, MG. (2006). The VPgPro protein of Turnip mosaic virus: in vitro inhibition of translation from a ribonuclease activity. *Virology* 351, 92-100.

[55] Khan, MA, Miyoshi, H, Gallie, DR, and Goss, DJ. (2008). Potyvirus genome-linked protein, VPg, directly affects wheat germ in vitro translation: interactions with translation initiation factors eIF4F and eIFiso4F. *J. Biol. Chem.* 283, 1340-9.

[56] Miyoshi, H, Okade, H, Muto, S, Suehiro, N, Nakashima, H, Tomoo, K, and Natsuaki, T. (2008). Turnip mosaic virus VPg interacts with Arabidopsis thaliana eIF(iso)4E and inhibits in vitro translation. *Biochimie* 90, 1427-34.

[57] Okade, H, Fujita, Y, Miyamoto, S, Tomoo, K, Muto, S, Miyoshi, H, Natsuaki, T, Rhoads, RE, and Ishida, T. (2009). Turnip mosaic virus genome-linked protein VPg binds C-terminal region of cap-bound initiation factor 4E orthologue without exhibiting host cellular specificity. *J. Biochem.* 145, 299-307.

[58] Carrington, JC, Freed, DD, and Leinicke, AJ. (1991). Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. *Plant Cell* 3, 953-62.

[59] Rajamäki, ML and Valkonen, JP. (2009). Control of Nuclear and Nucleolar Localization of Nuclear Inclusion Protein a of Picorna-Like Potato virus A in Nicotiana Species. *Plant Cell* 21, 2485-502.

[60] Lucy, AP, Guo, HS, Li, WX, and Ding, SW. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *EMBO J.* 19, 1672-80.

[61] Liu, K, Xia, Z, Zhang, Y, Wen, Y, Wang, D, Brandenburg, K, Harris, F, and Phoenix, DA. (2005). Interaction between the movement protein of barley yellow dwarf virus and the cell nuclear envelope: role of a putative amphiphilic alpha-helix at the N-terminus of the movement protein. *Biopolymers* 79, 86-96.

[62] Dennison, SR, Harris, F, Brandenburg, K, and Phoenix, DA. (2007). Characterization of the N-terminal segment used by the barley yellow dwarf virus movement protein to promote interaction with the nuclear membrane of host plant cells. *Peptides* 28, 2091-7.

[63] Hiscox, JA. (2003). The interaction of animal cytoplasmic RNA viruses with the nucleus to facilitate replication. *Virus Res.* 95, 13-22.

[64] Dougherty, WG and Hiebert, E. (1980). Translation of potyvirus RNA in a rabbit reticulocyte lysate: identification of nuclear inclusion proteins as products of tobacco etch virus RNA translation and cylindrical inclusion protein as a product of the potyvirus genome. *Virology* 104, 174-82.

[65] Himmelbach, A, Chapdelaine, Y, and Hohn, T. (1996). Interaction between cauliflower mosaic virus inclusion body protein and capsid protein: implications for viral assembly. *Virology* 217, 147-57.

[66] Nicolas, O, Dunnington, SW, Gotow, LF, Pirone, TP, and Hellmann, GM. (1997). Variations in the VPg protein allow a potyvirus to overcome va gene resistance in tobacco. *Virology* 237, 452-9.

[67] Dunoyer, P, Thomas, C, Harrison, S, Revers, F, and Maule, A. (2004). A cysteine-rich plant protein potentiates Potyvirus movement through an interaction with the virus genome-linked protein VPg. *J. Virol.* 78, 2301-9.

[68] Gao, Z, Johansen, E, Eyers, S, Thomas, CL, Noel Ellis, TH, and Maule, AJ. (2004). The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J.* 40, 376-85.

[69] Carrington, JC, Jensen, PE, and Schaad, MC. (1998). Genetic evidence for an essential role for potyvirus CI protein in cell-to-cell movement. *Plant J.* 14, 393-400.

[70] Roberts, IM, Wang, D, Findlay, K, and Maule, AJ. (1998). Ultrastructural and temporal observations of the potyvirus cylindrical inclusions (Cls) show that the Cl protein acts transiently in aiding virus movement. *Virology* 245, 173-81.

[71] Rodriguez-Cerezo, E, Findlay, K, Shaw, JG, Lomonossoff, GP, Qiu, SG, Linstead, P, Shanks, M, and Risco, C. (1997). The coat and cylindrical inclusion proteins of a potyvirus are associated with connections between plant cells. *Virology* 236, 296-306.

[72] Torrance, L, Andreev, IA, Gabrenaite-Verhovskaya, R, Cowan, G, Mäkinen, K, and Taliansky, ME. (2006). An unusual structure at one end of potato potyvirus particles. *J. Mol. Biol.* 357, 1-8.

[73] Gabrenaite-Verkhovskaya, R, Andreev, IA, Kalinina, NO, Torrance, L, Taliansky, ME, and Mäkinen, K. (2008). Cylindrical inclusion protein of potato virus A is associated with a subpopulation of particles isolated from infected plants. *J. Gen. Virol.* 89, 829-38.

[74] Rojas, MR, Zerbini, FM, Allison, RF, Gilbertson, RL, and Lucas, WJ. (1997). Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins. *Virology* 237, 283-95.

[75] Rajamäki, ML and Valkonen, JP. (2002). Viral genome-linked protein (VPg) controls accumulation and phloem-loading of a potyvirus in inoculated potato leaves. *Mol. Plant Microbe Interact.* 15, 138-49.

[76] Rajamäki, ML and Valkonen, JP. (2003). Localization of a potyvirus and the viral genome-linked protein in wild potato leaves at an early stage of systemic infection. *Mol. Plant Microbe Interact.* 16, 25-34.

[77] Ivanov, KI, Puustinen, P, Merits, A, Saarma, M, and Mäkinen, K. (2001). Phosphorylation down-regulates the RNA binding function of the coat protein of potato virus A. *J. Biol. Chem.* 276, 13530-40.

[78] Puustinen, P, Rajamäki, ML, Ivanov, KI, Valkonen, JP, and Mäkinen, K. (2002). Detection of the potyviral genome-linked protein VPg in virions and its phosphorylation by host kinases. *J. Virol.* 76, 12703-11.

[79] Hafrén, A and Mäkinen, K. (2008). Purification of viral genome-linked protein VPg from potato virus A-infected plants reveals several post-translationally modified forms of the protein. *J. Gen. Virol.* 89, 1509-18.

[80] Atabekov, JG, Rodionova, NP, Karpova, OV, Kozlovsky, SV, Novikov, VK, and Arkhipenko, MV. (2001). Translational activation of encapsidated potato virus X RNA by coat protein phosphorylation. *Virology* 286, 466-74.

[81] Karpova, OV, Rodionova, NP, Ivanov, KI, Kozlovsky, SV, Dorokhov, YL, and Atabekov, JG. (1999). Phosphorylation of tobacco mosaic virus movement protein abolishes its translation repressing ability. *Virology* 261, 20-4.

[82] Waigmann, E, Chen, MH, Bachmaier, R, Ghoshroy, S, and Citovsky, V. (2000). Regulation of plasmodesmal transport by phosphorylation of tobacco mosaic virus cell-to-cell movement protein. *EMBO J.* 19, 4875-84.

[83] Trutnyeva, K, Bachmaier, R, and Waigmann, E. (2005). Mimicking carboxyterminal phosphorylation differentially effects subcellular distribution and cell-to-cell movement of Tobacco mosaic virus movement protein. *Virology* 332, 563-77.

[84] Iakoucheva, LM, Radivojac, P, Brown, CJ, O'Connor, TR, Sikes, JG, Obradovic, Z, and Dunker, AK. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 32, 1037-49.

[85] Boggs, JM. (2006). Myelin basic protein: a multifunctional protein. *Cell Mol. Life Sci.* 63, 1945-61.

[86] Harauz, G, Ishiyama, N, Hill, CM, Bates, IR, Libich, DS, and Fares, C. (2004). Myelin basic protein-diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis. *Micron* 35, 503-42.

[87] Homchaudhuri, L, Polverini, E, Gao, W, Harauz, G, and Boggs, J. (2009). Influence of membrane surface charge and post-translational modifications to myelin basic protein on its ability to tether the Fyn-SH3 domain to a membrane in vitro. *Biochemistry*. In press.

[88] Sickmeier, M, Hamilton, JA, LeGall, T, Vacic, V, Cortese, MS, Tantos, A, Szabo, B, Tompa, P, Chen, J, Uversky, VN, Obradovic, Z, and Dunker, AK. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 35, D786-93.

[89] Romero, P, Obradovic, Z, Kissinger, CR, Villafranca, JE, Garner, E, Guilliot, S, and Dunker, AK. (1998). Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* , 437-48.

[90] Le Gall, T, Romero, PR, Cortese, MS, Uversky, VN, and Dunker, AK. (2007). Intrinsic disorder in the Protein Data Bank. *J. Biomol. Struct. Dyn.* 24, 325-42.

[91] Dunker, AK, Lawson, JD, Brown, CJ, Williams, RM, Romero, P, Oh, JS, Oldfield, CJ, Campen, AM, Ratliff, CM, Hipps, KW, Ausio, J, Nissen, MS, Reeves, R, Kang, C, Kissinger, CR, Bailey, RW, Griswold, MD, Chiu, W, Garner, EC, and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26-59.

[92] Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527-33.

[93] Dyson, HJ and Wright, PE. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197-208.

[94] Wright, PE and Dyson, HJ. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31-8.

[95] Uversky, VN, Gillespie, JR, and Fink, AL. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41, 415-27.

[96] Fink, AL. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15, 35-41.

[97] Dunker, AK, Oldfield, CJ, Meng, J, Romero, P, Yang, JY, Chen, JW, Vacic, V, Obradovic, Z, and Uversky, VN. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 9 Suppl 2, S1.

[98] Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.* 579, 3346-54.

[99] Uversky, VN. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11, 739-56.

[100] Bourhis, JM, Johansson, K, Receveur-Bréchot, V, Oldfield, CJ, Dunker, KA, Canard, B, and Longhi, S. (2004). The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res.* 99, 157-67.

[101] Duvignaud, JB, Savard, C, Fromentin, R, Majeau, N, Leclerc, D, and Gagné, SM. (2009). Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein. *Biochem. Biophys. Res. Commun.* 378, 27-31.

[102] Murray, CL, Marcotrigiano, J, and Rice, CM. (2008). Bovine viral diarrhea virus core is an intrinsically disordered protein that binds RNA. *J. Virol.* 82, 1294-304.

[103] Goh, GK, Dunker, AK, and Uversky, VN. (2009). Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virol. J.* 6, 69.

[104] Radivojac, P, Iakoucheva, LM, Oldfield, CJ, Obradovic, Z, Uversky, VN, and Dunker, AK. (2007). Intrinsic disorder and functional proteomics. *Biophys. J.* 92, 1439-56.

[105] Williams, RM, Obradovic, Z, Mathura, V, Braun, W, Garner, EC, Young, J, Takayama, S, Brown, CJ, and Dunker, AK. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.* , 89-100.

[106] Romero, P, Obradovic, Z, Li, X, Garner, EC, Brown, CJ, and Dunker, AK. (2001). Sequence complexity of disordered protein. *Proteins* 42, 38-48.

[107] Okorokov, AL and Orlova, EV. (2009). Structural biology of the p53 tumour suppressor. *Curr. Opin. Struct. Biol.* 19, 197-202.

[108] Haynes, C, Oldfield, CJ, Ji, F, Klitgord, N, Cusick, ME, Radivojac, P, Uversky, VN, Vidal, M, and Iakoucheva, LM. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol.* 2, e100.

[109] Dunker, AK, Cortese, MS, Romero, P, Iakoucheva, LM, and Uversky, VN. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129-48.

[110] Harauz, G, Ladizhansky, V, and Boggs, JM. (2009). Structural polymorphism and multifunctionality of myelin basic protein. *Biochemistry* 48, 8094-104.

[111] Oldfield, CJ, Meng, J, Yang, JY, Yang, MQ, Uversky, VN, and Dunker, AK. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9 Suppl 1, S1.

[112] Wright, PE and Dyson, HJ. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321-31.

[113] Xie, H, Vucetic, S, Iakoucheva, LM, Oldfield, CJ, Dunker, AK, Uversky, VN, and Obradovic, Z. (2007). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 6, 1882-98.

[114] Xie, H, Vucetic, S, Iakoucheva, LM, Oldfield, CJ, Dunker, AK, Obradovic, Z, and Uversky, VN. (2007). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* 6, 1917-32.

[115] Vucetic, S, Xie, H, Iakoucheva, LM, Oldfield, CJ, Dunker, AK, Obradovic, Z, and Uversky, VN. (2007). Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J. Proteome Res.* 6, 1899-916.

[116] Yoon, MK, Shin, J, Choi, G, and Choi, BS. (2006). Intrinsically unstructured N-terminal domain of bZIP transcription factor HY5. *Proteins* 65, 856-66.

[117] Liu, J, Perumal, NB, Oldfield, CJ, Su, EW, Uversky, VN, and Dunker, AK. (2006). Intrinsic disorder in transcription factors. *Biochemistry* 45, 6873-88.

[118] Johnson, LN and Lewis, RJ. (2001). Structural basis for control by phosphorylation. *Chem. Rev.* 101, 2209-42.

[119] Ferreira, ME, Hermann, S, Prochasson, P, Workman, JL, Berndt, KD, and Wright, AP. (2005). Mechanism of transcription factor recruitment by acidic activators. *J. Biol. Chem.* 280, 21779-84.

[120] Kleerekoper, QK and Putkey, JA. (2009). PEP-19, an intrinsically disordered regulator of calmodulin signaling. *J. Biol. Chem.* 284, 7455-64.

[121] Oldfield, CJ, Cheng, Y, Cortese, MS, Romero, P, Uversky, VN, and Dunker, AK. (2005). Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44, 12454-70.

[122] Shoemaker, BA, Portman, JJ, and Wolynes, PG. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8868-73.

[123] Huang, Y and Liu, Z. (2009). Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism. *J. Mol. Biol.* .

[124] Turjanski, AG, Gutkind, JS, Best, RB, and Hummer, G. (2008). Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput. Biol.* 4, e1000060.

[125] Espinoza-Fonseca, LM. (2009). Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem. Biophys. Res. Commun.* 382, 479-82.

[126] Yang, JY, Yang, MQ, Dunker, AK, Deng, Y, and Huang, X. (2008). Investigation of transmembrane proteins using a computational approach. *BMC Genomics* 9 Suppl 1, S7.

[127] Schwartz, M, Chen, J, Lee, WM, Janda, M, and Ahlquist, P. (2004). Alternate, virus-induced membrane rearrangements support positive-strand RNA virus genome replication. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11263-8.

[128] Davidson, WS, Jonas, A, Clayton, DF, and George, JM. (1998). Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* 273, 9443-9.

[129] Volles, MJ and Lansbury, PT,Jr. (2002). Vesicle permeabilization by protofibrillar alpha-synuclein is sensitive to Parkinson's disease-linked mutations and occurs by a pore-like mechanism. *Biochemistry* 41, 4595-602.

[130] Kjaer, L, Giehm, L, Heimburg, T, and Otzen, D. (2009). The influence of vesicle size and composition on alpha-synuclein structure and stability. *Biophys. J.* 96, 2857-70.

[131] Boettcher, JM, Hartman, KL, Ladror, DT, Qi, Z, Woods, WS, George, JM, and Rienstra, CM. (2008). Membrane-induced folding of the cAMP-regulated phosphoprotein endosulfine-alpha. *Biochemistry* 47, 12357-64.

[132] Koag, MC, Fenton, RD, Wilkens, S, and Close, TJ. (2003). The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiol.* 131, 309-16.

[133] Koag, MC, Wilkens, S, Fenton, RD, Resnik, J, Vo, E, and Close, TJ. (2009). The K-segment of maize DHN1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. *Plant Physiol.* 150, 1503-14.

[134] Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19, 23-30.

[135] Receveur-Bréchot, V, Bourhis, JM, Uversky, VN, Canard, B, and Longhi, S. (2006). Assessing protein disorder and induced folding. *Proteins* 62, 24-45.

[136] Bracken, C, Iakoucheva, LM, Romero, PR, and Dunker, AK. (2004). Combining prediction, computation and experiment for the characterization of protein disorder. *Curr. Opin. Struct. Biol.* 14, 570-6.

[137] Ferron, F, Longhi, S, Canard, B, and Karlin, D. (2006). A practical overview of protein disorder prediction methods. *Proteins* 65, 1-14.

[138] Greenfield, NJ. (2004). Analysis of circular dichroism data. *Methods Enzymol.* 383, 282-317.

[139] Adler, AJ, Greenfield, NJ, and Fasman, GD. (1973). Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol.* 27, 675-735.

[140] Kelly, SM and Price, NC. (2000). The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* 1, 349-84.

[141] Kelly, SM, Jess, TJ, and Price, NC. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta* 1751, 119-39.

[142] Andrade, MA, Chacon, P, Merelo, JJ, and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* 6, 383-90.

[143] Sreerama, N and Woody, RW. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287, 252-60.

[144] Dyson, HJ and Wright, PE. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 104, 3607-22.

[145] Selenko, P and Wagner, G. (2007). Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* 158, 244-53.

[146] Serber, Z, Corsini, L, Durst, F, and Dotsch, V. (2005). In-cell NMR spectroscopy. *Methods Enzymol.* 394, 17-41.

[147] McNulty, BC, Young, GB, and Pielak, GJ. (2006). Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *J. Mol. Biol.* 355, 893-7.

[148] Mucsi, Z, Hudecz, F, Hollosi, M, Tompa, P, and Friedrich, P. (2003). Binding-induced folding transitions in calpastatin subdomains A and C. *Protein Sci.* 12, 2327-36.

[149] Toke, O, Banoczi, Z, Tarkanyi, G, Friedrich, P, and Hudecz, F. (2009). Folding transitions in calpain activator peptides studied by solution NMR spectroscopy. *J. Pept. Sci.* 15, 404-10.

[150] Sugase, K, Dyson, HJ, and Wright, PE. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447, 1021-5.

[151] Eftink, MR and Ghiron, CA. (1981). Fluorescence quenching studies with proteins. *Anal. Biochem.* 114, 199-227.

[152] Matyus, L, Szollosi, J, and Jenei, A. (2006). Steady-state fluorescence quenching applications for studying protein structure and dynamics. *J. Photochem. Photobiol. B.* 83, 223-36.

[153] Khaymina, SS, Kenney, JM, Schroeter, MM, and Chalovich, JM. (2007). Fesselin is a natively unfolded protein. *J. Proteome Res.* 6, 3648-54.

[154] Semisotnov, GV, Rodionova, NA, Razgulyaev, OI, Uversky, VN, Gripas', AF, and Gilmanshin, RI. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* 31, 119-28.

[155] Bailey, RW, Dunker, AK, Brown, CJ, Garner, EC, and Griswold, MD. (2001). Clusterin, a binding protein with a molten globule-like region. *Biochemistry* 40, 11828-40.

[156] Kuznetsova, IM, Turoverov, KK, Dunker, AK, and Uversky, VN. Analysis of folded, partially folded and misfolded proteins with fluorescent dyes. In *Methods in Protein Structure and Stability Analysis: Luminescence Spectroscopy and Circular Dichroism* (Uversky, V.N. and Permyakov, E.A., eds.) pp. 73-104. Nova Science Publishers, Inc., Hauppauge, NY, USA.

[157] Iakoucheva, LM, Kimzey, AL, Masselon, CD, Smith, RD, Dunker, AK, and Ackerman, EJ. (2001). Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci.* 10, 1353-62.

[158] Hubbard, SJ. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* 1382, 191-206.

[159] Fontana, A, de Laureto, PP, Spolaore, B, Frare, E, Picotti, P, and Zambonin, M. (2004). Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* 51, 299-321.

[160] Fontana, A, Zambonin, M, Polverino de Laureto, P, De Filippis, V, Clementi, A, and Scaramella, E. (1997). Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol.* 266, 223-30.

[161] Kim, YJ, Kim, YA, Park, N, Son, HS, Kim, KS, and Hahn, JH. (2005). Structural characterization of the molten globule state of apomyoglobin by limited proteolysis and HPLC-mass spectrometry. *Biochemistry* 44, 7490-6.

[162] Romero, P, Obradovic, Z, Kissinger, CR, Villafranca, JE, and Dunker, AK. (1997). Identifying disordered regions in proteins from amino acid sequences. *Proceedings of IEEE International Conference on Neural Networks* 1, 90-5.

[163] Li, X, Romero, P, Rani, M, Dunker, AK, and Obradovic, Z. (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* 10, 30-40.

[164] Peng, K, Radivojac, P, Vucetic, S, Dunker, AK, and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7, 208.

[165] Obradovic, Z, Peng, K, Vucetic, S, Radivojac, P, Brown, CJ, and Dunker, AK. (2003). Predicting intrinsic disorder from amino acid sequence. *Proteins* 53 Suppl 6, 566-72.

[166] Ward, JJ, McGuffin, LJ, Bryson, K, Buxton, BF, and Jones, DT. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138-9.

[167] Vacic, V, Uversky, VN, Dunker, AK, and Lonardi, S. (2007). Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 8, 211.

[168] Clertant, P and Cuzin, F. (1982). Covalent affinity labeling by periodate-oxidized [alpha-32P]ATP of the large-T proteins of polyoma and SV40 viruses. *J. Biol. Chem.* 257, 6300-5.

[169] Merits, A, Guo, D, and Saarma, M. (1998). VPg, coat protein and five non-structural proteins of potato A potyvirus bind RNA in a sequence-unspecific manner. *J. Gen. Virol.* 79 (Pt 12), 3123-7.

[170] Tellez, AB, Crowder, S, Spagnolo, JF, Thompson, AA, Peersen, OB, Brutlag, DL, and Kirkegaard, K. (2006). Nucleotide channel of RNA-dependent RNA polymerase used for intermolecular uridylylation of protein primer. *J. Mol. Biol.* 357, 665-75.

[171] Rajamäki, ML and Valkonen, JP. (1999). The 6K2 protein and the VPg of potato virus A are determinants of systemic infection in Nicandra physaloides. *Mol. Plant Microbe Interact.* 12, 1074-81.

[172] Kentsis, A and Sosnick, TR. (1998). Trifluoroethanol promotes helix formation by destabilizing backbone exposure: desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding. *Biochemistry* 37, 14613-22.

[173] Herce, HD and Garcia, AE. (2007). Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20805-10.

[174] Shai, Y. (1999). Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim. Biophys. Acta* 1462, 55-70.

[175] Brogden, KA. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* 3, 238-50.

[176] Pouny, Y, Rapaport, D, Mor, A, Nicolas, P, and Shai, Y. (1992). Interaction of antimicrobial dermaseptin and its fluorescently labeled analogues with phospholipid membranes. *Biochemistry* 31, 12416-23.

[177] Russ, WP and Engelman, DM. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* 296, 911-9.

[178] Senes, A, Engel, DE, and DeGrado, WF. (2004). Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr. Opin. Struct. Biol.* 14, 465-79.

[179] Curran, AR and Engelman, DM. (2003). Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* 13, 412-7.

[180] Hildebrand, PW, Lorenzen, S, Goede, A, and Preissner, R. (2006). Analysis and prediction of helix-helix interactions in membrane channels and transporters. *Proteins* 64, 253-62.

[181] DeGrado, WF, Gratkowski, H, and Lear, JD. (2003). How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci.* 12, 647-65.

[182] Spencer, RH and Rees, DC. (2002). The alpha-helix and the organization and gating of channels. *Annu. Rev. Biophys. Biomol. Struct.* 31, 207-33.

[183] Gonzalez, ME and Carrasco, L. (2003). Viroporins. *FEBS Lett.* 552, 28-34.

[184] Mackenzie, J. (2005). Wrapping things up about virus RNA replication. *Traffic* 6, 967-77.

[185] Schaad, MC, Jensen, PE, and Carrington, JC. (1997). Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *EMBO J.* 16, 4049-59.

[186] Restrepo-Hartwig, MA and Carrington, JC. (1994). The tobacco etch potyvirus 6-kilodalton protein is membrane associated and involved in viral replication. *J. Virol.* 68, 2388-97.

[187] Wei, T and Wang, A. (2008). Biogenesis of cytoplasmic membranous vesicles for plant potyvirus replication occurs at endoplasmic reticulum exit sites in a COPI- and COPII-dependent manner. *J. Virol.* 82, 12252-64.

[188] Dufresne, PJ, Thivierge, K, Cotton, S, Beauchemin, C, Ide, C, Ubalijoro, E, Laliberté, JF, and Fortin, MG. (2008). Heat shock 70 protein interaction with Turnip mosaic virus RNA-dependent RNA polymerase within virus-induced membrane vesicles. *Virology* 374, 217-27.

[189] Thivierge, K, Cotton, S, Dufresne, PJ, Mathieu, I, Beauchemin, C, Ide, C, Fortin, MG, and Laliberté, JF. (2008). Eukaryotic elongation factor 1A interacts with Turnip mosaic virus RNA-dependent RNA polymerase and VPg-Pro in virus-induced vesicles. *Virology* 377, 216-25.

[190] Peränen, J, Laakkonen, P, Hyvönen, M, and Kääriäinen, L. (1995). The alphavirus replicase protein nsP1 is membrane-associated and has affinity to endocytic organelles. *Virology* 208, 610-20.

[191] Ahola, T, Lampio, A, Auvinen, P, and Kääriäinen, L. (1999). Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity. *EMBO J.* 18, 3164-72.

[192] Verchot, J, Koonin, EV, and Carrington, JC. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* 185, 527-35.

[193] Rajamäki, ML, Kelloniemi, J, Alminaite, A, Kekarainen, T, Rabenstein, F, and Valkonen, JP. (2005). A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology* 342, 88-101.

[194] Valli, A, Martín-Hernández, AM, López-Moya, JJ, and García, JA. (2006). RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro. *J. Virol.* 80, 10055-63.

[195] Wang, RY, Powell, G, Hardie, J, and Pirone, TP. (1998). Role of the helper component in vector-specific transmission of potyviruses. *J. Gen. Virol.* 79 (Pt 6), 1519-24.

[196] Carrington, JC and Herndon, KL. (1992). Characterization of the potyviral HC-pro autoproteolytic cleavage site. *Virology* 187, 308-15.

[197] Sáenz, P, Salvador, B, Simón-Mateo, C, Kasschau, KD, Carrington, JC, and García, JA. (2002). Host-specific involvement of the HC protein in the long-distance movement of potyviruses. *J. Virol.* 76, 1922-31.

[198] Ballut, L, Drucker, M, Pugnière, M, Cambon, F, Blanc, S, Roquet, F, Candresse, T, Schmid, HP, Nicolas, P, Gall, OL, and Badaoui, S. (2005). HcPro, a multifunctional protein encoded by a plant RNA virus, targets the 20S proteasome and affects its enzymic activities. *J. Gen. Virol.* 86, 2595-603.

[199] Choi, IR, Horken, KM, Stenger, DC, and French, R. (2005). An internal RNA element in the P3 cistron of Wheat streak mosaic virus revealed by synonymous mutations that affect both movement and replication. *J. Gen. Virol.* 86, 2605-14.

[200] Merits, A, Guo, D, Jarvekulg, L, and Saarma, M. (1999). Biochemical and genetic evidence for interactions between potato A potyvirus-encoded proteins P1 and P3 and proteins of the putative replication complex. *Virology* 263, 15-22.

[201] Hong, XY, Chen, J, Shi, YH, and Chen, JP. (2007). The '6K1' protein of a strain of Soybean mosaic virus localizes to the cell periphery. *Arch. Virol.* 152, 1547-51.

[202] Lain, S, Riechmann, JL, and García, JA. (1990). RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. *Nucleic Acids Res.* 18, 7003-6.

[203] Spetz, C and Valkonen, JP. (2004). Potyviral 6K2 protein long-distance movement and symptom-induction functions are independent and host-specific. *Mol. Plant Microbe Interact.* 17, 502-10.

[204] Schaad, MC, Lellis, AD, and Carrington, JC. (1997). VPg of tobacco etch potyvirus is a host genotype-specific determinant for long-distance movement. *J. Virol.* 71, 8624-31.

[205] Carrington, JC and Dougherty, WG. (1987). Small Nuclear Inclusion Protein Encoded by a Plant Potyvirus Genome Is a Protease. *J. Virol.* 61, 2540-8.

[206] Anindya, R and Savithri, HS. (2004). Potyviral NIa proteinase, a proteinase with novel deoxyribonuclease activity. *J. Biol. Chem.* 279, 32159-69.

[207] Domier, LL, Shaw, JG, and Rhoads, RE. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology* 158, 20-7.

[208] Hong, Y and Hunt, AG. (1996). RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology* 226, 146-51.

[209] Atreya, CD, Raccah, B, and Pirone, TP. (1990). A point mutation in the coat protein abolishes aphid transmissibility of a potyvirus. *Virology* 178, 161-5.

[210] Dolja, VV, Haldeman-Cahill, R, Montgomery, AE, Vandenbosch, KA, and Carrington, JC. (1995). Capsid protein determinants involved in cell-to-cell and long distance movement of tobacco etch potyvirus. *Virology* 206, 1007-16.

[211] Mahajan, S, Dolja, VV, and Carrington, JC. (1996). Roles of the sequence encoding tobacco etch virus capsid protein in genome amplification: requirements for the translation process and a cis-active element. *J. Virol.* 70, 4370-9.

[212] Ward, JJ, Sodhi, JS, McGuffin, LJ, Buxton, BF, and Jones, DT. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337, 635-45.

[213] Linding, R, Jensen, LJ, Diella, F, Bork, P, Gibson, TJ, and Russell, RB. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* 11, 1453-9.

[214] Prilusky, J, Felder, CE, Zeev-Ben-Mordehai, T, Rydberg, EH, Man, O, Beckmann, JS, Silman, I, and Sussman, JL. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21, 3435-8.

[215] Dosztányi, Z, Csizmok, V, Tompa, P, and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433-4.

[216] Gaboriauda, C, Bissery, V, Benchetrit, T, Mornon, J, P. (1987). Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences *FEBS Lett.* 224, 149-155.